НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

ФАКУЛЬТЕТ ЕЛЕКТРОНІКИ

КАФЕДРА ПРОМИСЛОВОЇ ЕЛЕКТРОНІКИ

«На правах рукопису»

УДК 621.314

До захисту допущено

Завідувач кафедри

_____ Ю.С. Ямненко

(підпис) (ініціали, прізвище)

“___”_____20__ р.

МАГІСТЕРСЬКА ДИСЕРТАЦІЯ

зі спеціальності 171 Електроніка _

(код та назва напряму підготовки або спеціальності)

на тему Пристрій розпізнавання голосу для керування кліматом в приміщенні _____

Виконав: студент _6_ курсу, групи __ДС-61м

___Береза Владислав Миколайович_____ _____

(прізвище, ім’я, по батькові) (підпис)

Науковий керівник ___к.т.н., доцент Хохлов Ю.В. _____

(посада, вчене звання, науковий ступінь, прізвище та ініціали) (підпис)

Рецензент_____ _____

(посада, вчене звання, науковий ступінь, прізвище та ініціали) (підпис)

Засвідчую, що у цьому дипломному проекті немає запозичень з праць інших авторів без відповідних посилань.

Студент _____

(підпис)

Київ – 2018 року

АНОТАЦІЯ

Останнім часом спостерігається значне зростання інтересу до технологій, пов'язаних з розпізнаванням мови. Завдання управління пристроями за допомогою голосових команд, інтерактивні платформи які надають інформацію після запиту в більш природній формі - за допомогою голосу, все це знаходить широкого застосування в сучасному світі. Багато задач виникає при бажанні взаємодіяти за допомогою голосу з мобільними пристроями. Наприклад, введення голосових команд для отримання інформації з Інтернету, прокладання маршруту руху, запуск програм користувача, диктування тексту. Останнім часом з'явилася можливість управління домашньою, офісною технікою за допомогою електронних пристроїв голосовими командами.

Передумовою розвитку голосових технологій є значне збільшення обчислювальних можливостей, обсягу пам'яті при значному зменшенні габаритів комп'ютерних систем. Слід також відзначити розвиток математичних методів, що дозволяють виконати необхідну обробку аудіо сигналу шляхом виділення з нього інформативних ознак.

Для прикладу, широко використовується дискретне перетворення Фур'є, яке відоме з теорії цифрової обробки сигналів. Подальша обробка виконується з використанням акустичної моделі, яка ставить у відповідність виділених параметрах конкретні звуки (фонеми). Слід зазначити, що в статті розглядається можливість використання методу динамічного програмування і метод нейронних рекурентних мереж.

Розглянуто взаємодію людини і домашніх приладів яка реалізовується на базі Arduino, за допомогою голосових команд. Тут передбачається, що обробка аудіо сигналу, побудова текстового рядка на основі виголошеній фрази виконується спеціальними бібліотеками (класами, методами), шо покращує розпізнавання голосу в порівнянні з існуючими методами.

SUMMARY

Recently, there has been a significant increase in interest in technologies related to speech recognition. The tasks of controlling devices with the help of voice commands, interactive platforms, provide information after the request in a more natural form - with the help of voice. Many tasks arise when you want to manage with voice with mobile devices. For example, the introduction of voice commands to obtain information from the Internet, the provide route, the launch of user programs, the dictation of the text. Recently, it became possible to manage home, office equipment using electronic devices with voice commands.

The prerequisite for the development of speech technologies is a significant increase in computing capabilities, memory capacity with a significant reduction in the size of computer systems. It should also be noted the development of mathematical methods that make it possible to perform the necessary processing of an audio signal by isolating informative features from it.

For example, a discrete Fourier transform is widely used, which is known from the theory of digital signal processing. Further processing is performed using an acoustic model, which assigns specific sounds to the selected parameters (phonemes). It should be noted that the article considers the possibility of using the dynamic programming method and the method of neural recurrent networks.

The interaction between humans and home devices, that implemented of platform Arduino, with the help of voice commands is considered. Here it is assumed that the processing of an audio signal, the construction of a text string based on the pronounced phrase is performed by special libraries (classes, methods), which improves voice recognition in comparison with existing methods.

ЗМІСТ

ВСТУП

Сучасні інформаційні технології прогресивно розвиваються у всіх аспектах та напрямках, тож цей процес впливає на наші повсякденні речі: чи то автомобіль який має власний бортовий комп’ютер для збору та аналізу даних, що виводить інформацію яка в подальшому сприяє покращенню процесів обслуговування автомобіля. Інший приклад - це «розумний» будильник який за допомогою додаткових датчиків аналізує фази сну і підбирає оптимальний час для пробудження користувача.

Достатньо перспективним є додавання голосового інтерфейсу до комп’ютеризованих систем керування промисловими та побутовими пристроями. Одною з актуальних проблем, яка вирішується при розробці таких систем керування, є проблема недостатньої точності розпізнавання голосових команд. Вдосконалення ведеться в напрямку підвищення надійності, незалежності від індивідуальних характеристик голосу, зниження негативного впливу фонового шуму на якість розпізнавання.

Попит на такі системи зараз обмежений з-за недостатньо високих параметрів і проблем пов'язаних з темою голосової безпеки.

Після аналізу попиту на використання голосового управління в побутових та промислових системах було прийнято рішення спроектувати пристрій який володіє достатніми параметрами сприйняття і обробки мови, та має прийнятні витрати на розробку і створення.

Важливим предметом досліджень є знаходження способів підвищення точності, позбавлення від шумів та інших факторів які впливають на процес розпізнавання голосу.

Для покращення точності розпізнавання голосу в пристрої стали використовувати глибинні нейронні мережі (ГНМ), які в останні роки неодноразово показували суттєві результати в процесах прогнозування, класифікації, розпізнавання образів, рукописного тексту та мовлення. Тому

використання ГНМ та їх модифікації у задачах розпізнавання мови є актуальною задачею сьогодення.[2]

Метою дослідження є вдосконалення точності розпізнавання голосу в порівнянні з використанням існуючих методів нейронних мереж та запропонованими в дисертації.

Об'єктом дослідження у даній магістерській дисертації є процес розпізнання голосу за допомогою запропонованого методу для керування кліматом в приміщенні.

Предмет дослідження – пристрій розпізнавання голосу для керування кліматом в приміщенні на основі використання рекурентних нейронних мереж.

Методи дослідження. Поставлені у роботі задачі вирішувалися шляхом проведення теоретичних та експериментальних досліджень. При аналізі можливості збільшення показників точності в розпізнаванні голосу використано основні положення математичного аналізу та комп'ютерних технологій. При дослідженні роботи схеми керування кліматом використано сучасні методи та програмні засоби Arduino IDE.

Практичне значення одержаних результатів полягає в наступному:

Розроблений робочий алгоритм для розпізнавання голосу з використанням рекурентних нейронних мереж який показав вищі показниками точності в порівнянні з існуючими методами.

Побудовано імітаційну модель схеми керуванням кліматом з використанням сучасної платформи Arduino.

Особистий внесок магістранта

Наукові положення та результати викладені в дисертації автором особисто.

В друкованих працях, опублікованих у співавторстві, особисто здобувачу належить: в [15] – дані досліджень використання РНМ в порівнянні з існуючими методами.

Публікації:

Основні наукові положення дисертації представлено в науковому журналі «Альманах науки».

РОЗДІЛ І. СУЧАСНЕ ГОЛОСОВЕ УПРАВЛІННЯ

1.1 Голосове та інтелектуальне управління.

Інтелектуальне управління і штучний інтелект (ШІ) – це окрема галузь науки, яка займається проектуванням комп'ютерних систем, які спроможні виконувати задачі на які не були спеціально запрограмовані. Такі системи спроможні навчатися аналогічного до того, як навчається людський мозок.

ШІ пов'язаний зі схожим завданням, застосування комп'ютерів для того щоб зрозуміти людський інтелект, але не завжди за основу беруть біологічно правдоподібні методи.

Історія штучного інтелекту в плані нової науки починається в середині XX століття. До цього часу вже сформувалася безліч передумов для зародження, активно тривали дискусії серед філософів про людську природу і процеси пізнання розуму особистості, психологи і нейрофізіологи розробили ряд теорій стосовно роботи людського мислення і мозку, математики та економісти ставили собі питання оптимальних розрахунків і представлення знань про світ в формалізованому вигляді, зародився фундамент математичної теорії обчислень теорії алгоритмів і були створені перші комп'ютери.

Можливості винайдених машин в плані швидкості обчислень виявилися більше людських, тому в науковому співтоваристві зародилося питання на тему меж можливостей комп'ютерів і розвитку мислення машин до рівня людини.

Логічний підхід до створення штучного інтелекту і його систем заснований на моделюванні міркувань, теоретичною основою є логіка.

Логічний підхід може бути проілюстрований застосуванням для цих цілей системи логічного програмування Prolog. Програми, записані на мові Prolog, представляють набори фактів і правил логічного висновку, без точного задання алгоритму як послідовності дій, що призводять до необхідного результату.

Останній підхід, що розвивається з початку 1990-х років, називається агентно-орієнтованим підходом, або підходом, заснованим на використанні інтелектуальних (раціональних) агентів. Відповідно до цього підходу, інтелект - це обчислювальна частина яка здатна досягати поставлених перед інтелектуальної машиною цілей. Сама така машина буде інтелектуальним агентом, що сприймає навколишній світ за допомогою датчиків, і здатною впливати на об'єкти в навколишньому середовищі за допомогою виконавчих механізмів. Цей підхід акцентує увагу на тих методах і алгоритмах, які допоможуть інтелектуальному агенту впливати на об’єкти в навколишньому середовищу за допомогою допоміжних механізмів.

Цей підхід акцентує увагу на тих методах і алгоритмах, які допоможуть інтелектуальному агенту виживати в навколишньому середовищі при виконанні його завдання. В цьому підході значно ретельніше вивчаються алгоритми пошуку шляху та прийняття рішень.

Стосовно до біологічних завдань агентно-орієнтований підхід називають також індивідуально-орієнтованим підходом. Він заснований на використанні клітинних автоматів. Такий підхід поєднує символьний і логічний підходи і дозволяє створювати математичні моделі складних систем.

Аналізуючи історію ШІ, можна виділити такий великий напрямок як моделювання міркувань. Довгі роки розвиток цієї науки рухалось саме цим шляхом, і тепер це одна з найрозвиненіших областей в сучасному ШІ. Моделювання міркувань має на увазі створення символьних систем, на вході яких поставлена визначена задача, а на виході потрібно її рішення. Як правило, завдання вже формалізоване, тобто переведене в математичну форму, але не має алгоритму рішення, або він занадто складний, трудомісткий і т.п. В цей напрямок входять: доведення теорем, прийняття рішень, планування і диспетчеризація, прогнозування.

Важливим напрямком є обробка природної мови, в рамках якого проводиться аналіз можливостей розуміння, обробки і генерації текстів на «людській» мові. В рамках цього напрямку ставиться мета такої обробки

природної мови, яка була б в змозі отримати знання самостійно, читаючи існуючий текст, доступний через Інтернет. Деякі прямі застосування обробки природної мови включають інформаційний пошук (в тому числі, глибокий аналіз тексту) і машинний переклад.

Напрямок інженерії знань об'єднує завдання отримання знань з простої інформації, їх систематизації та використання. Цей напрямок історично пов'язаний зі створенням експертних систем - програм, що використовують спеціалізовані бази знань для отримання достовірних висновків з рішення проблеми.

Отримання знань з даних - одна з базових проблем інтелектуального аналізу даних. Існують різні підходи до вирішення цієї проблеми, в тому числі - на основі технології з використанням нейронних мереж, що використовують процедури вербалізації нейронних мереж.

Регресійний аналіз використовується, в виявленні безперервної функції, на підставі якої можна було б прогнозувати вихід. При навчанні агент отримує позитивні сигнали за правильні відповіді і негативні за погані. Вони можуть бути проаналізовані з точки зору теорії рішень, використовуючи такі поняття як корисність. Математичний аналіз машинних алгоритмів вивчення - це розділ теоретичної інформатики, відомий як обчислювальна теорія навчання. До області машинного навчання відноситься великий клас задач на розпізнавання образів. Наприклад, це розпізнавання символів, мови, рукописного тексту, аналіз текстів. Багато завдання успішно вирішуються за допомогою біологічного моделювання. Варто тільки згадати комп'ютерний зір, який також пов'язаний з робототехнікою.

Галузі робототехніки і штучного інтелекту тісно пов'язані один з одним. Інтегрування цих двох наук, створення інтелектуальних роботів складають ще один напрямок ШІ. Інтелектуальна складова потрібна роботам, щоб маніпулювати об'єктами, виконувати навігацію з проблемами локалізації (визначати місцезнаходження, вивчати найближчі області) і планувати рух (як дістатися до мети). Прикладом таких приладів робототехніки можуть служити

автоматичні машини які здатні відтворювати деякі звички людини. Природа людської творчості ще менш вивчена, ніж природа інтелекту. Проте, ця область існує, і тут поставлені проблеми написання комп'ютером музики, літературних творів, художня творчість. Створення реалістичних образів широко використовується в кіно і індустрії ігор.

Окремо виділяється вивчення проблем технічної творчості систем штучного інтелекту.

Додавання такої можливості до будь-якої інтелектуальної системи дозволяє досить наочно продемонструвати, що саме система розуміє і як це сприймає. Додавання шуму замість відсутньої інформації або фільтрація шуму наявними в системі знаннями, виробляє з абстрактних знань конкретні образи, що легко сприймаються людиною, особливо це корисно для інтуїтивних і малоцінних знань, перевірка яких в формальному вигляді вимагає значних розумових зусиль.

Деякі з найвідоміших ШІ-систем:

а) Watson - перспективна розробка IBM, здатна сприймати людську мову і виробляти імовірнісний пошук, із застосуванням великої кількості алгоритмів.

б) Розпізнавання мови, ViaVoice система яка обслуговує клієнтів.

Банки застосовують системи штучного інтелекту (СШІ) в страховій діяльності, при грі на біржі і управлінні власністю. Методи розпізнавання образів широко використовують при оптичному і акустичному розпізнаванні, в тому числі тексту й мови, медичній діагностиці, спам-фільтрах, в системах ППО (визначення цілей), а також для забезпечення ряду інших завдань національної безпеки.

У комп'ютерних науках проблеми штучного інтелекту розглядаються з позицій проектування експертних систем та баз знань. Під базами знань розуміється сукупність даних і правил виведення, що допускають логічний висновок і осмислену обробку інформації. В цілому дослідження проблем штучного інтелекту в комп'ютерних науках спрямовані на створення, розвиток

і експлуатацію інтелектуальних інформаційних систем, а питання підготовки користувачів і розробників таких систем вирішуються фахівцями інформаційних технологій [1].

1.2 Поняття, призначення і види голосового управління

1.2.1 Поняття голосового управління

Голосове управління з'явилося в суспільстві відносно недавно. Воно має на увазі під собою перетворення мови в цифрову інформацію. Перший пристрій подібного роду з'явилося в 1952 році, він міг розпізнавати вимовлені людиною цифри.

В основі голосового управління лежить досить проста схема роботи – передавач -> приймач -> вихід. В якості передавача голосової команди виступає окремий модуль прийому голосових команд, або їм може виявитися смартфон з встановленим на нього певним програмним забезпеченням, також комп'ютер при наявності мікрофона і програми.

Приймачем виступає пристрій голосового управління. Крім прийому інформації завдання пристрою перетворити його в цифровий сигнал або інформацію для виконання команди або набору тексту. Приймачем також може бути передавач, якщо це програма на комп'ютер або смартфон.

Також голосове управління використовується для набору тексту, відкриття вікон, браузерів, пошукових програм та іншого.

Від приймача цифровий сигнал через виходи йде на підключення до входів пристрою, наприклад на телевізор, комп'ютер та світильник. Таким чином подаючи 1 на один з входів, підключений до нього пристрій включиться, подаючи сигнал 0 він вимкнеться.

На рис. 1.1 схематично зображено роботу голосового управління.

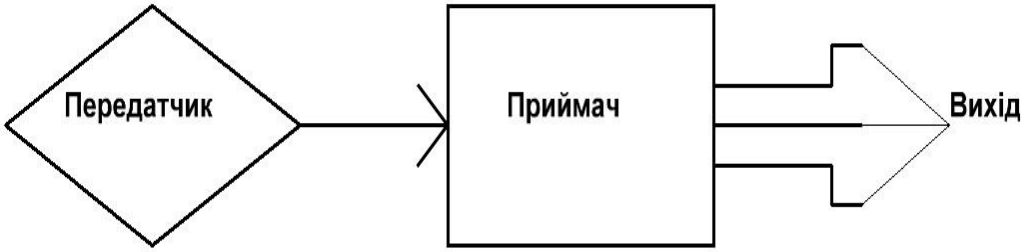


Рис. 1.1 Схема роботи голосового управління

Комерційний розвиток голосове управління отримало на початку дев'яностих років. У той час потужність обчислювальної техніки стрімко зростала, що дозволяло крім створення окремих пристроїв управління голосом створити програми для персональних комп'ютерів і надалі для мобільних телефонів, планшетів та інших пристроїв.

Сучасні пристрої голосового управління мають високу точність обробки і перетворення мови в цифрову інформацію і можуть сприймати те чи інше слово для виконання команди. Для цього в лістингу пристрою потрібно позначити команду як ключове слово. На даний момент програмне забезпечення вже дозволяє ввести, змінити ключове слово без втручання додаткових програм для прошивки, тобто силами самої програми.

В наші дні голосове управління може бути безконтактним, тобто зазвичай для приведення пристрою в режим прийому голосових команд потрібно натиснути певну кнопку в програмі мобільного телефону, персонального комп'ютера, модуля прийому голосових команд. Однак останні версії програм голосового управління мають можливість завдання ключового слова для переходу пристрою в режим прийому голосу або команди і подальшого задання вже робочих команд.

1.2.2 Призначення пристроїв розпізнавання мови

Призначення пристроїв розпізнавання мови і голосових команд полягає в полегшенні доступу до техніки в будинку, машині, виробництві, роботі з

документами і іншого. Все більше застосування голосове управління отримує в бізнесі, медицині, роботі в офісі. До розвитку голосового управління в сфері бізнесу можна віднести телефонію: автоматична обробка вхідних і вихідних дзвінків шляхом створення голосових систем самообслуговування зокрема для: отримання довідкової інформації та консультування, замовлення послуг/ товарів, зміни параметрів діючих послуг, проведення опитувань, анкетування, збору інформації, інформування та будь-які інші сценарії. Також в сучасному автомобілебудуванні розпізнавання мовлення відіграє важливу роль, зокрема управління опціями в салоні автомобіля, такими як навігація, мультимедіа та інше. Розпізнавання мови набула широкого застосування в мобільних додатках: голосове управління смартфона, закладене спочатку або доступне для скачування, активується за допомогою певної послідовності дій. Також голосовий пошук користується популярністю, а саме такі програми як, наприклад, Яндекс навігація або GoogleNow для персональних комп'ютерів і така команда як «Окей, Google» для смартфонів.

Велику поширеність розпізнавання мови отримує у людей з обмеженими можливостями або тимчасовими травмами. Так, наприклад людина з травмою рук за допомогою голосових пристроїв може набрати повідомлення і відправити його на певну адресу або номер, а працівник на виробництві може перемикати режим роботи механізму завданням ключового слова для переходу в режим прийому голосової команди і далі - задати саму голосову команду, після чого механізм перейде в інший режим роботи.

На основі голосового управління створена не одна система «Розумний будинок». У даній системі модуль голосового управління підключається до техніки в будинку, таким як лампи, комп'ютер, телевізор, двері, вікна та інше. Для модуля є своє програмне забезпечення, яке підтримується смартфонами, комп'ютером. Встановлюючи програму голосового управління даним модулем і з'єднуючись з ним за допомогою Bluetooth, Wi-Fi можна дистанційно керувати майже всією технікою в будинку. Закрити вікно, штори, двері, включити-вимкнути світло, телевізор, комп'ютер, праску - все це можливо

робити на відстані за допомогою Bluetooth, або з будь-якої точки світу за допомогою Wi-Fi.

У наш час також розвивається робототехніка і створення штучного інтелекту, для яких діалог з людиною просто необхідний. Голосове управління відіграє головну роль при взаємодії людини зі штучним інтелектом [5].

1.2.3 Види голосового управління.

Є кілька видів голосового управління. Одні діляться за способом з'єднання, інші по функціоналу. Однак, так, чи інакше, вони всі є пристроями розпізнавання мови.

За типом з'єднання можна виділити з'єднання по кабелю, а також Bluetooth і Wi-Fi з'єднання. Наприклад, для передачі мови може використовуватися окремий модуль прийому голосових команд, який з'єднаний з модулем обробки і перетворення мови в сигнал через кабель, або знаходиться в його складі як єдиний модуль.

Але слід зазначити, що в наші дні найбільш поширена схема використання програмного забезпечення на смартфон, комп'ютер і передача мови для обробки через Bluetooth або Internet по Wi-Fi. Bluetooth з'єднання не вимагає підключення до мережі Internet і трохи простіше у використанні, ніж Wi-Fi, однак дальність передачі сигналу відносно невелика.

Передача інформації через Internet по Wi-Fi зі смартфона вимагає наявності підключення, в зв'язку з чим будуть витрати на підключення, але суттєву перевагу даного методу полягає в тому, що управляти пристроєм можна з будь-якої точки земної кулі.

В основному при голосовому управлінні використовують тільки поодинокі слова, так як при введенні тексту доведеться витримувати певну паузу між словами і надрукувати цей текст буде швидше. Однак люди з обмеженими можливостями або травмами активно використовують голосовий набір тексту.

Сучасне програмне забезпечення для голосового управління підтримує широкий спектр операційних систем на комп'ютер, смартфон, планшет і т.п. Обов'язковою умовою при голосовому управлінні є наявність мікрофона у пристрою, через який планується передавати голосову інформацію [3].

1.3 Архітектура і ознаки пристроїв голосового управління

1.3.1 Архітектура пристроїв розпізнавання мови

Пристрої перетворення мови в цифровий сигнал або інформацію мають власну архітектуру.

Елемент очищення від шуму і елемент виділення корисного сигналу, частіше поєднані в єдиний модуль.

Для будь-якого окремого звуку будується статистична модель, яка описує проголошення даного звуку в мові.

Голосова бібліотека збирається на основі окремої мови і її складність прямо пропорційна обраній мові. Наприклад, зібрати голосову бібліотеку англійською значно простіше, ніж українською. Голосова бібліотека відіграє важливу роль при голосовому наборі речень або словосполучень.

Декодер - елемент розпізнавання мови, поєднує отриману голосову інформацію і на основі моделі і бібліотеки визначає межі кожного слова, що дозволяє розпізнати злиту мову.

1.3.2 Ознаки в пристроях розпізнавання мови.

Основні поняття, які характеризують мову людини, пов'язані з формою, розмірами, динамікою зміни голосу і описують емоційний стан людини, тож їх можна розділити на чотири групи об'єктивних ознак, що дозволяють розрізняти голосові зразки: спектрально-часові, кепстральні, амплітудно-частотні та ознаки нелінійної динаміки.

Спектральні ознаки

а) Середнє значення спектра аналізованого голосового сигналу.

б) Нормалізовані середні значення спектра.

в) Відносний час перебування сигналу в смугах спектра.

г) Нормалізований час перебування сигналу в смугах спектра.

д) Медіанне значення спектра мови в смугах.

е) Відносна потужність спектра мови в смугах.

ж) Варіація огинаючих спектрів мови.

і) Нормалізовані величини варіації огинаючих спектрів мови.

п) Коефіцієнти кросскорреляції спектральних огинаючих між смугами спектра.

Часові ознаки

а) Тривалість сегмента, фонеми.

б) Висота сегмента.

в) Коефіцієнт форми сегмента.

Спектрально-часові ознаки характеризують голосовий сигнал в його фізико-математичної суті виходячи з наявності компонентів трьох видів:

а) періодичних (тональних) ділянок звукової хвилі;

б) неперіодичних ділянок звукової хвилі (шумових, вибухових);

в) ділянок, що не містять голосових пауз.

Спектрально-часові ознаки дозволяють відображати своєрідність форми тимчасового ряду і спектра голосових імпульсів у різних осіб і особливості фільтруючих функцій їх голосових трактів.

Характеризують особливості голосового потоку, які пов'язані з динамікою перебудови артикуляційних органів мови людини, і є інтегральними характеристиками голосового потоку, що відображають своєрідність взаємозв'язку або синхронності руху артикуляційних органів мовця.

Кепстральні ознаки

а) Мел-частотні кепстральні коефіцієнти.

б) Основні лінійні коефіцієнти з поправкою на нерівномірність чутливості людського вуха.

в) Коефіцієнти потужності частоти реєстрації.

г) Коефіцієнти спектра лінійного передбачення.

д) Коефіцієнти кепстра лінійного передбачення.

Більшість сучасних автоматичних систем розпізнавання мови зосереджують зусилля на отриманні частотної характеристики голосового тракту людини, відкидаючи при цьому характеристики сигналу збудження. Це пояснено тим, що коефіцієнти першої моделі забезпечують кращу роздільність звуків.

Для відділення сигналу збудження від сигналу голосового тракту вдаються до кепстральних аналізу.

Амплітудно-частотні ознаки

а) Інтенсивність, амплітуда.

б) Енергія.

в) Частота основного тону (чот).

г) Фомантні частоти.

д) Джіттер - тремтіння частотної модуляції основного тону.

е) Шіммер- амплітудна модуляція на основному тоні.

Амплітудно-частотні ознаки дозволяють отримувати оцінки, значення яких можуть змінюватися в залежності від параметрів дискретного перетворення Фур'є (виду і ширини вікна), а також при незначних зрушеннях вікна по вибірці. Голосовий сигнал акустично є поширюваний в повітряному середовищі, складні за своєю структурою звукові коливання які характеризуються відносно їх частоти (числа коливань в секунду), інтенсивності (амплітуди коливань) і тривалості. Амплітудно-частотні ознаки несуть необхідну і достатню інформацію для людини по мовному сигналу при

мінімальному часу сприйняття. Але застосування цих ознак не дозволяє в повній мірі використовувати їх як інструмент ідентифікації емоційно-забарвленої мови.

Ознаки нелінійної динаміки

а) Відображення Пуанкаре.

б) рекурентний графік.

в) Максимальний показник Ляпунова - емоційний стан людини, якому відповідає певна геометрія аттрактора.

г) Фазовий портрет (аттрактор).

д) Розмірність Каплана-Йорк - кількісна міра емоційного стану людини, від «спокою» до «гніву».

Для ознак нелінійної динаміки голосовий сигнал розглядається як скалярна величина, яка спостерігається в системі голосового тракту людини.

Процес мовостворення можна вважати нелінійним і аналізувати його методами нелінійної динаміки. Завдання нелінійної динаміки полягає в знаходженні і докладному дослідженні базових математичних моделей і реальних систем, які виходять з найбільш типових пропозицій про властивості окремих елементів, що складають систему, і закони взаємодії між ними. В даний час методи нелінійної динаміки базуються на фундаментальній математичній теорії, в основі якої лежить теорема Такенса, яка підводить сувору математичну основу під ідеї нелінійної авторегресії і доводить можливість відновлення фазового портрета аттрактора по тимчасовому ряду або по одній його координаті (під аттрактором розуміють безліч точок або підпростір в фазовому просторі, до якого наближається фазова траєкторія після загасання перехідних процесів). Оцінки характеристик сигналу з відновлених голосових траєкторій використовуються в побудові нелінійних детермінованих фазово-просторових моделей спостережуваного тимчасового ряду. Виявлені відмінності в формі аттракторів можна використовувати для

діагностичних правил і ознак, що дозволяють розпізнати і правильно ідентифікувати різні емоції в емоційно пофарбованому мовному сигналі.

Відповідно до теорії мовостворення мова являє собою акустичну хвилю, яка випромінюється системою органів: легкими, бронхами і трахеєю, а потім перетворюється в голосовому тракті. Якщо припустити, що джерела збудження і форма голосового тракту відносно незалежні, мовний апарат людини можна представити у вигляді сукупності генераторів тональних сигналів, шумів та фільтрів. Схематично це представлено на рис. 1.2., де:

1. Генератор імпульсної послідовності (тонів);
2. Генератор випадкових чисел (шумів);
3. Коефіцієнти цифрового фільтра (параметри голосового тракту);
4. Нестаціонарний цифровий фільтр.

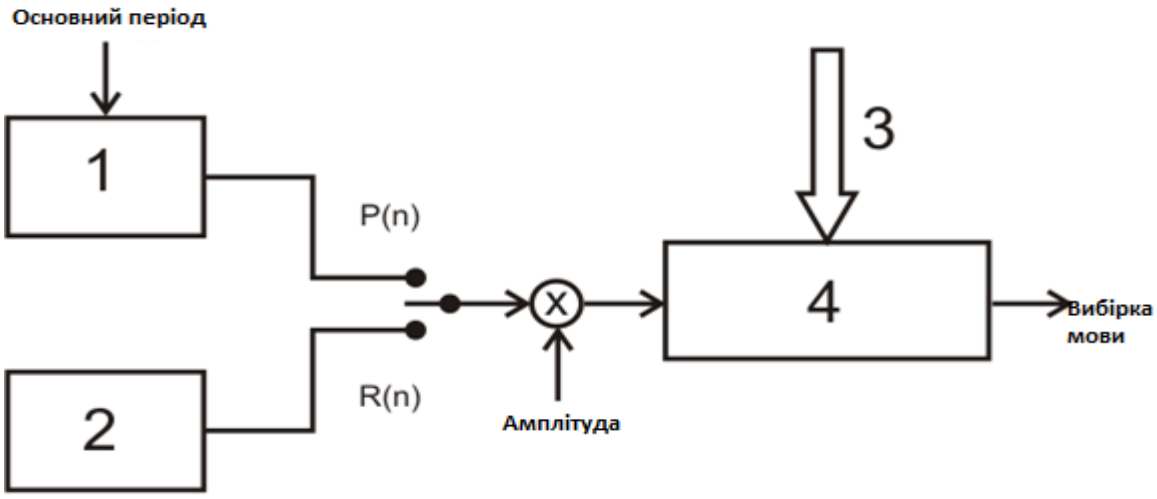


Рис.1.2. Схема голосового апарату людини

1.3.3 Параметри якості мови й основні поняття

Параметри якості мови:

а) Складова розбірливість мови.

б) Фразова розбірливість мови.

в) Якість мови в порівнянні з еталоном.

г) Якість мови в реальних умовах роботи.

Основні поняття:

а) Чіткість голосу - кількість правильно прийнятих елементів мови (звуків, складів, слів, фраз), виражене у відсотках від загального числа переданих елементів.

б) Якість мови - характеризує суб'єктивну оцінку звучання мови в випробуваної системі передачі мови.

в) Нормальний темп мови - проголошення промови з такою швидкістю, при якій середня тривалість контрольної фрази дорівнює 2,4/с.

г) Прискорений темп мови - проголошення промови з такою швидкістю, при якій середня тривалість контрольної фрази дорівнює 1,5-1,6 с.

д) Розпізнання голосу мовця - дає можливість слухачам порівнювати звучання голосу, з конкретною особою, відомим слухачеві раніше.

е) Смислова розбірливість - вказує на ступінь правильного відтворення інформаційного змісту промови.

ж) Інтегральна якість - характеризує більш загальне враження слухача від прийнятої мови.

Якість мови і добротність є основним фактором в діалозі між людиною і голосовим інтерфейсом, оскільки точно зрозуміти який текст або команда буде виголошена завдання першорядної важливості, яка допоможе уникнути непорозумінь і як наслідок виконання невірної команди або введення неправильного тексту [7].

1.4 Синтез мови в голосовому управлінні.

1.4.1 Синтез мови.

Синтез мови - відновлення за параметрами голосового сигналу, формування сигналу мовлення з друкованого тексту. Синтез мови можна застосувати всюди там, де одержувачем є людина. Якість синтезатора мови залежить від його схожості з людським голосом і зрозумілою мовою. Це

дозволяє людям зі сліпотою слухати письмові роботи на домашньому комп'ютері.

Синтез мови використовують в наступних випадках:

а) Інформаційно - довідкові системи.

б) Видача інформації про поточні технологічні процеси.

в) Створення музики.

Існує кілька способів синтезу мови:

а) Параметричний синтез.

б) Компіляційний синтез.

в) Синтез за правилами.

г) Предметно-орієнтований синтез.

Параметричний синтез мови є кінцевою операцією в кодуючих системах, де мовний сигнал представляється набором невеликого числа безперервно змінюваних параметрів. Параметричний синтез доцільно застосовувати в тих випадках, коли набір повідомлень обмежений і змінюється не дуже часто. Перевагою такого способу є можливість записати голос для будь-якої мови і будь-якого диктора. Якість параметричного синтезу може бути дуже високою (в залежності від ступеня стиснення інформації в параметричному представленні). Однак параметричний синтез не може застосовуватися для довільних, заздалегідь не заданих повідомлень.

Компіляційний синтез зводиться до складання повідомлення з попередньо записаного словника вихідних елементів синтезу. Розмір елементів синтезу не менш слова. Очевидно, що зміст синтезованих повідомлень фіксується обсягом словника. Як правило, число одиниць словника не перевищує декількох сотень слів. Основна проблема в компілятивному синтезі - обсяги пам'яті для зберігання словника.

У зв'язку з цим використовуються різноманітні методи стиснення/кодування мовного сигналу. Компілятивний синтез має широке практичне застосування. У західних країнах різноманітні пристрої (від військових літаків до побутових пристроїв) оснащуються системами голосової

відповіді. В Україні системи мовної відповіді знаходять все більше застосування в повсякденному житті, наприклад, в довідкових службах операторів стільникового зв'язку при отриманні інформації про стан рахунку абонента.

Повний синтез мови за правилами (або синтез за друкованим текстом) забезпечує управління всіма параметрами мовного сигналу і, таким чином, може генерувати мову по заздалегідь невідомому тексту. В цьому випадку параметри, отримані при аналізі мовного сигналу, зберігаються в пам'яті так само, як і правила для з'єднання звуків в слова і фрази.

Синтез реалізується шляхом моделювання мовного тракту, застосування аналогової або цифрової техніки. При цьому в процесі синтезування значення параметрів і правила з'єднання фонем вводять послідовно через певний часовий інтервал, наприклад 5-10 мс. Метод синтезу мови за друкованим текстом (синтез за правилами) базується на запрограмованому знанні акустичних і лінгвістичних обмежень і не використовує елементи людської мови. У системах, заснованих на цьому способі синтезу, виділяється два підходи. Перший підхід спрямований на побудову моделі мови похідної від системи людини, він відомий під назвою артикуляторного синтезу. Другий підхід - формантний синтез за правилами. Чіткість і натуральність таких синтезаторів може бути доведена до величин, порівнянних з характеристиками природної мови. Синтез мови за правилами з використанням попередньо збережених відрізків природної мови - це різновид синтезу мови за правилами, яка набула поширення в зв'язку з появою можливостей маніпулювання голосовим сигналом в оцифрованої формі. Залежно від розміру вихідних елементів синтезу виділяються такі види синтезу:

а) мікросегментний (мікрохвильовий);

б) аллофонічний;

в) дифонний;

г) напівскладовий;

д) складовий;

ж) синтез з одиниць довільного розміру.

Зазвичай в якості таких елементів використовуються напівскладові сегменти, що містять половину приголосного і половину приєднаного до нього голосного. При цьому можна синтезувати мову по заздалегідь не заданому тексту, але важко керувати інтонаційними характеристиками. Якість такого синтезу не відповідає якості природної мови, оскільки на кордонах зшивання дифонів часто виникають спотворення. Компіляція мови із заздалегідь записаних словоформ також не вирішує проблеми високоякісного синтезу довільних повідомлень, оскільки акустичні і просодичні (тривалість і інтонація) характеристики слів змінюються в залежності від типу фрази і місця слова у фразі. Це положення не змінюється навіть при використанні великих обсягів пам'яті для зберігання словоформ.

Предметно-орієнтований синтез компілює слова, записані заздалегідь, а також фрази для створення повних голосових повідомлень. Він використовується в додатках, де різноманіття текстів системи буде обмежене певною темою/областю, наприклад оголошення про відправлення поїздів і прогнози погоди. Ця технологія проста у використанні і досить довго застосовувалася в комерційних цілях: її так само застосовували при виготовленні електронних приладів, таких як годинник, що виголошує інформацію. Природність звучання цих систем потенційно може бути високою завдяки тому, що різноманіття видів пропозицій обмежена і близько з відповідністю інтонацією вихідних записів. А так як ці системи обмежені вибором слів і фраз в базі даних, вони в подальшому не можуть мати широке поширення в сферах діяльності людини, лише тому, що здатні синтезувати комбінації слів і фраз, на які вони були запрограмовані.

1.5 Оцифрування звуку

1.5.1 Оцифрування сигналу

Оцифрування звуку грає головну роль в системах ГК (рис.1.3). Цифровий звук є аналоговим звуковим сигналом, представлений у вигляді числових значень -амплітуди звуку. Оцифрування звуку включає в себе два процеси:

а) дискретизація;

б) квантування амплітуди.

Процес дискретизації по часу - процес отримання значень сигналу, який перетворюється з певним часовим кроком дискретизації. Кількість замірів величини сигналу, що здійснюються в одну секунду, називають частотою дискретизації або частотою вибірки. Чим менше крок, тим більша їх кількість і тим більш точне уявлення про сигнал буде отримано. Це підтверджується теоремою Котельникова. Згідно з теоремою, аналоговий сигнал з обмеженим спектром точно описується дискретною послідовністю значень його амплітуди, якщо ці значення беруться з частотою, яка удвічі перевищує найвищу частоту спектра сигналу. На рис. 1.3. наведено принцип роботи АЦП.

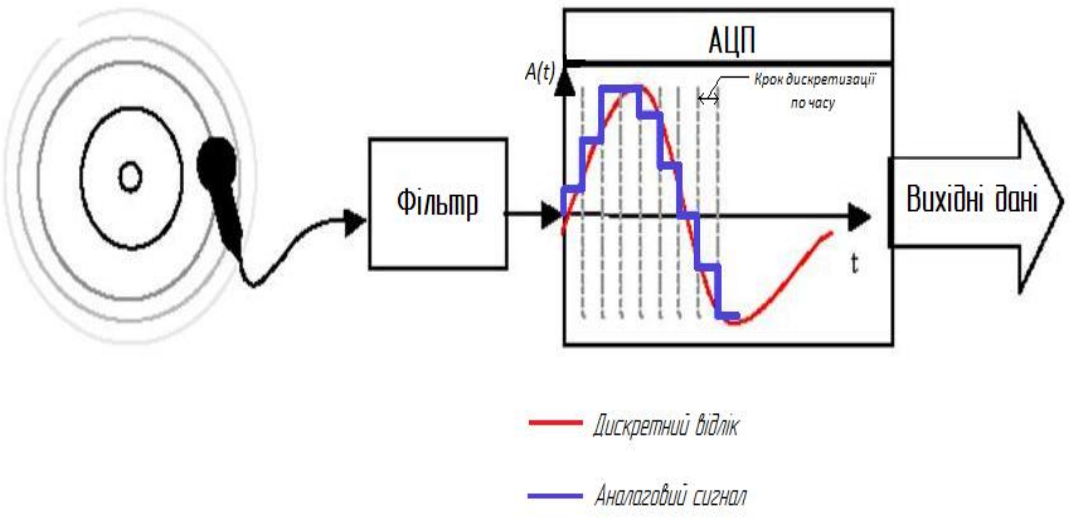


Рис.1.3 Робота АЦП

Тобто, аналоговий сигнал, в якому частота спектра пряма то сигнал може бути точно представлений послідовністю дискретних значень амплітуди. На практиці це означає, що для того, щоб оцифрований сигнал містив інформацію про всьому діапазоні чутних частот вихідного аналогового сигналу (0 - 20 кГц) необхідно, щоб вибране значення частоти дискретизації становило не менше 40 кГц. Кількість замірів амплітуди в секунду називають частотою дискретизації. Основні труднощі оцифрування полягають в неможливості записати виміряні значення сигналу з ідеальною точністю. На рис. 1.4 докладно представлена схема оцифрування голосу.

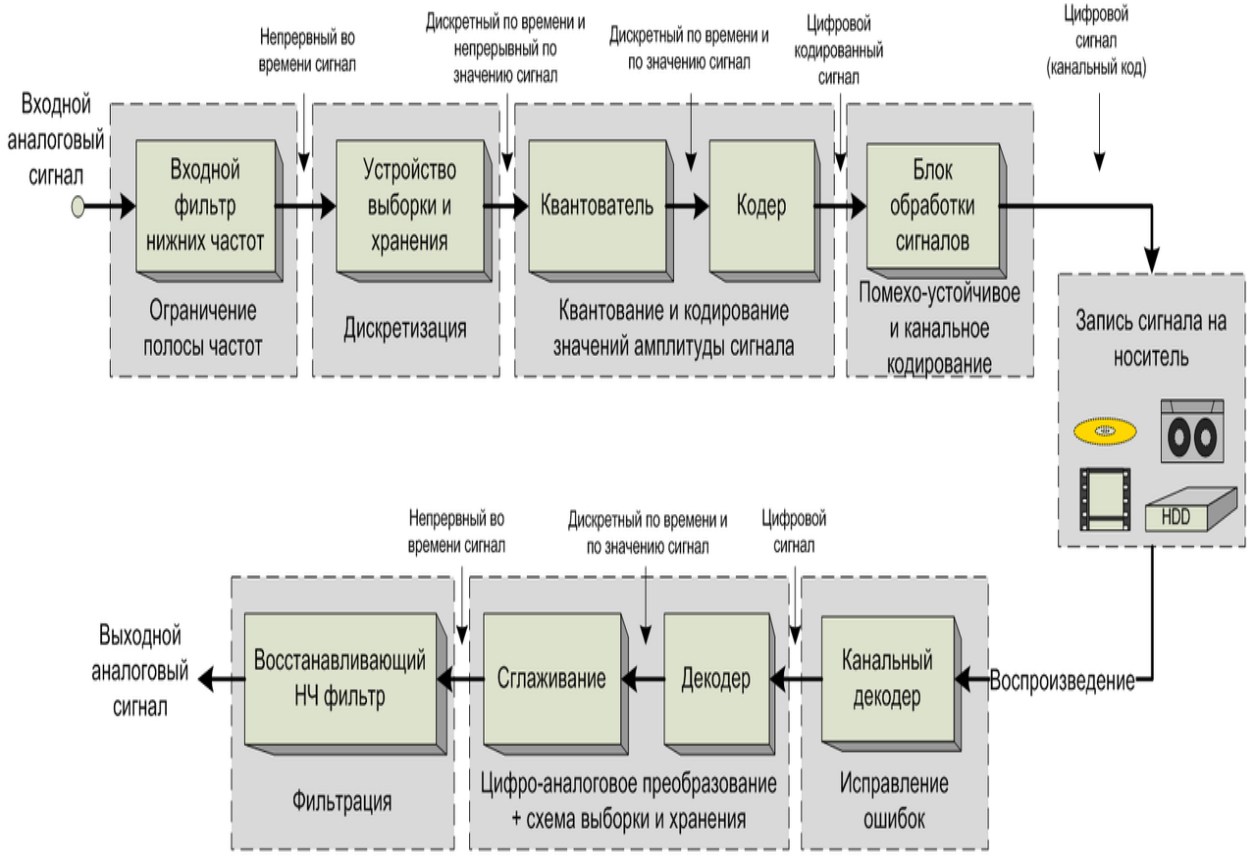


Рис.1.4 - Схема оцифрування голосу

1.5.2 Шуми

У цифровому звуці можна виділити два основних джерела шумів.

Джитер (тремтіння)

Це випадкові відхилення сигналу, як правило, виникають через нестабільність частоти тактового генератора або різної швидкості поширення різних частотних складових одного сигналу. Ця проблема може виникнути на стадії оцифрування. Це відбувається через різні відстані між вертикальними лініями (рис.1.5).

У цифровому звукозаписі слід використовувати високоякісні кварцові генератори з джерелами живлення, які мають малі пульсації і шуми. Застосування повністю цифрових студій також дозволяє звести вплив джитера до мінімуму. Такою студією може бути і персональний комп'ютер зі звуковою платою, що має хороший АЦП, в разі зберігання, редагування та відтворення звуку тільки в цифровому вигляді.

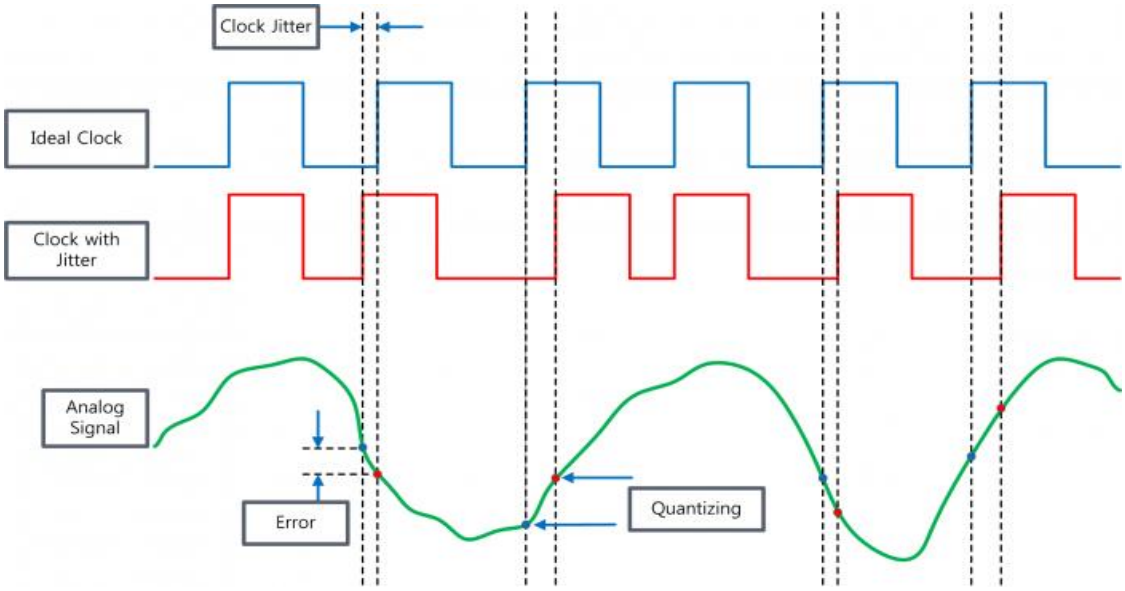


Рис.1.5. Явище джитер при оцифруванні сигналу.

Алиасінг

При оцифрування можлива ситуація, при якій в цифровому сигналі можуть з'явитися частотні складові, яких не було в оригінальному сигналі. Дана помилка отримала назву Aliasing. Цей ефект безпосередньо пов'язаний з частотою дискретизації, а точніше - з частотою Найквіста. Найпростіше зрозуміти, як це відбувається, розглянувши рис1.6.:

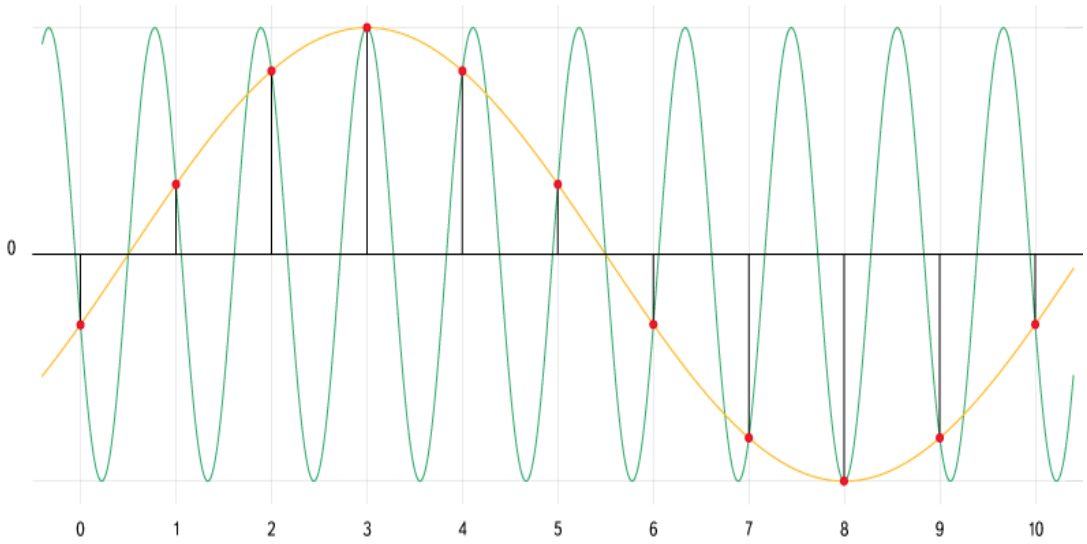


Рис.1.6 Ефект Aliasing

1.6 Аналіз ринку систем голосового управління

На зростання ринку розпізнавання голосу і управління з його допомогою впливають деякі фактори і розвиток йде неоднозначно. У зв'язку з відносно недавнім появою систем ГК, важливим фактором є безпека, так як в деяких сферах діяльності важлива кодування голосу диктора, щоб уникнути зломів і доступу до голосового управління сторонній особі. В наші дні безпеку голосового керування не настільки розвинена і новітні розробки в області систем розпізнавання мови націлені на її поліпшення.

Великим недоліком в даній області розробок є їх висока вартість, що істотно сповільнює процес розвитку подібних систем. Також в даний час варто відзначити неможливість придушення зовнішніх шумів, яка ускладнює процес більш точного розпізнавання голосу і не дає системам ГК твердо розташуватися в світовій економіці і помітно вплинути на неї, оскільки дані системи займають вкрай маленький відсоток на технічному ринку.

Важливу роль, як на ринку, так і в розвитку систем РГ і ГУ грає голосова біометрія, яка використовується на секретних об'єктах, таких як військові бази або наукові лабораторії. Загальною проблемою на ринку подібних систем є

відносно низькі показники точності, які зараз активно поліпшуються за допомогою голосової біометрії і інших систем розпізнавання мови.

Спочатку, слово «біометрія» зустрічалося тільки в медичній теорії. Проте, стали зростати потреби в безпеці з використанням біометричних технологій серед підприємств і державних установ. Використання біометричних технологій - один з ключових чинників на світовому ринку розпізнавання мови.

Розпізнавання голосу використовується перевірки автентичності людини, так як голос кожної людини індивідуальний. Це забезпечить високий рівень точності і безпеки. Розпізнавання голосу має велике значення в фінансових інститутах, таких як банк, а так само на підприємствах в сфері охорони здоров'я. В даний час сегмент розпізнавання мови становить 3,5% від частки технологій біометрії на світовому ринку, але це частка має постійне зростання. Також низька вартість біометричних пристроїв збільшує попит з боку малого і середнього бізнесу.

Військові відомства в більшості країн використовують вкрай обмежені зони для того, щоб запобігти проникненню зловмисників. Для забезпечення секретності і безпеки в цій зоні, військові використовують системи розпізнавання голосу. Ці системи допомагають військовим установам виявляти наявність несанкціонованих проникнень в захищену зону. Система містить базу даних голосів військовослужбовців і державних чиновників, які мають допуск до захищеної території. Ці люди ідентифікуються системою розпізнавання голосу, тим самим запобігається допуск людей, чиїх голосів немає в базі даних системи.

На додаток можна сказати, що військові використовують голосові команди для керування літаком. Крім того, військові відомства використовують розпізнавання мови і систему Voice-to-text для комунікації з громадянами в інших країнах. Наприклад, американські військові активно використовують системи розпізнавання мови в їх операціях в Іраку і

Афганістані. Таким чином, існує високий попит на розпізнавання мови і голосу для військових цілей.

Ефект від проблем, що стоять перед ринком, як очікується, повинен звести нанівець наявність різних тенденцій, які з'являються на ринку. Однією з таких тенденція є збільшення попиту на розпізнавання мови на мобільних пристроях.

Усвідомлюючи величезний потенціал мобільних пристроїв, виробники на світовому ринку розпізнавання голосу розвивають інноваційні додатки, специфічні для роботи на мобільних пристроях. Це один з майбутніх рушійних чинників. Зростаючий попит на голосову аутентифікацію мобільного банкінгу є ще однією позитивною тенденцією на ринку розпізнавання голосу. Деякі з основних тенденцій на світовому ринку розпізнавання голосу:

а) Збільшення попиту на програми розпізнавання мови на мобільних пристроях.

б) Зростання попиту на послуги голосової аутентифікації для мобільного банкінгу.

в) Інтеграція голосової верифікації і розпізнавання мови.

г) Збільшення злиттів і поглинань.

Найбільша частка попиту з систем РГ і ГК доводиться на мобільні додатки, в основному для смартфонів, сучасних магнітол в автомобілебудуванні і системи розумного будинку, які суттєво полегшують доступ до будь-якої техніки в житлових секторах. Активно дані системи застосовуються і в сферах військової діяльності, медицині і промисловому виробництві.

Зростаюче число правил дорожнього руху, що забороняють використання мобільних пристроїв під час водіння автомобіля, збільшило попит на додатки розпізнавання мови. Країни, в яких були накладені суворі обмеження: Австралія, Філіппіни, США, Великобританія, Індія і Чилі. У США більш ніж в 13 штатах, не дивлячись на введення положення про використання

мобільних пристроїв, дозволено використовувати гучний зв'язок під час водіння.

Отже, покупці все частіше вибирають мобільні пристрої, оснащені додатками розпізнавання мови, які зможуть допомогти їм отримати доступ до пристрою без необхідності відволікатися на сам пристрій. З метою задоволення зростаючого попиту на додатки розпізнавання мови в мобільних пристроях, виробники збільшили кількість науково-дослідних і дослідно-конструкторських робіт для того, щоб розвинути мовні команди опцій для мобільного пристрою. В результаті, велика кількість додатків розпізнавання мови були включені в мобільний пристрій, наприклад, управління музичним плей листом, зчитування адреси, зчитування імені абонента, голосові СМС повідомлення і т.д. Серед компаній на ринку займаються комерційним поширенням систем РГ і ГК найбільш конкурентноздатними будуть ті, чиї системи будуть найбільш точними і комбінованими (наприклад, правильне розпізнавання мови буде супроводжуватися і підтверджуватися відео розпізнаванням). Отже, при неоднозначних тенденцій ринку голосового управління і розпізнавання мови ринок все ж розвивається. У період з 2012 по 2016 рік ринок подібних систем збільшився в 1,5 - 2 рази в залежності від того, на що спрямована та, чи інша система в окремо. На рис. 1.7 показані тенденції розвитку ГУ на ринку [9].

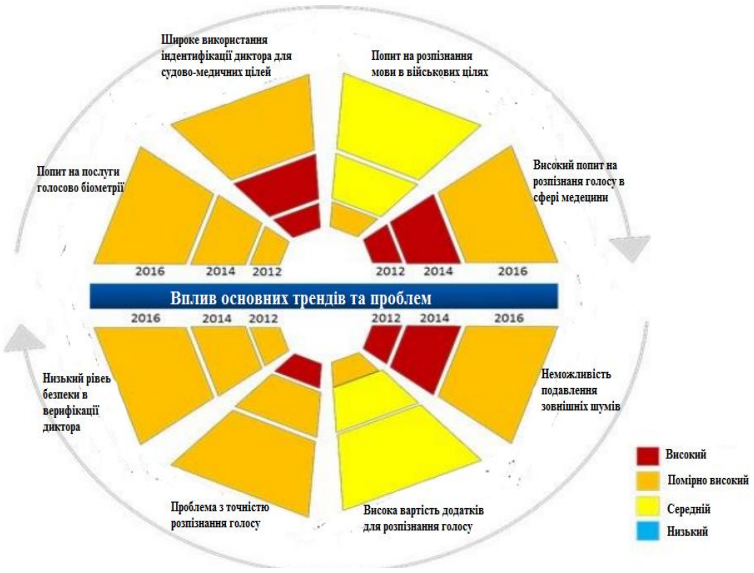


Рис.1.7 Актуальність пристроїв для розпізнавання голосу

1.7 Задача керування кліматом

Створення потрібного мікроклімату в приміщенні досягається узгодженою роботою цілого ряду приладів (рис. 1.8).

Рис.1.8 Управління кліматом в системі «Розумний будинок»

Це можуть бути електричні та газові котли, радіатори, системи «теплої» підлоги, кондиціонери, зволожувачі повітря, система вентиляції. Управління цими приладами в ручному режимі створює масу незручностей. Необхідно кожен раз налаштовувати їх роботу в відповідно до погодних умов, часу доби, уподобаннями господаря і т.п. Тим більше, це складно, якщо в кожному приміщенні необхідно створити свій мікроклімат. У Розумному будинку управління кліматом здійснюється за допомогою персонального комп'ютера, мобільного телефону або єдиної панелі управління. При цьому знадобиться всього кілька натискань для створення персонального мікроклімату для кожного приміщення.

Основна роль, при створенні клімату в приміщенні, відводиться джерел тепла. Це можуть бути батареї опалення, конвертори, теплі підлоги, в деяких випадках, кондиціонер. Господар вибере найбільш економічно вигідний, в даний час доби, джерело тепла.

Так, наприклад, якщо встановлені два котла: електричний і газовий, то з настанням ночі господар може переключити з газового опалення на електричне, яке в нічний час значно дешевше.

Наявність спеціальних датчиків температури дозволяє реагувати на зміну погодних умов за вікном. У теплу погоду є можливість самостійно відключити «зайве» опалення, економлячи тим самим витрати електрики.

1.8 Аналіз останніх досліджень і результатів.

Традиційні системи розпізнавання мови були засновані на математичному апараті прихованих марковських моделей. Російський математик Андрій Марков, в честь якого названо моделі, при дослідженні задач обробки літературних текстів на початку XX століття оцінював імовірність появи кожної літери в тексті в залежності від її контексту. Для спрощення обчислень він припустив, що ці ймовірності залежать тільки від однієї попередньої літери - Марківська властивість. Виявилося, що оцінки ймовірностей переходу від однієї букви до іншої за різними фрагментами одного тексту практично ідентичні. Надалі з'ясувалася унікальність параметрів марковської моделі (ланцюга Маркова) для кожного учасника, що дозволило застосувати їх в задачах визначення авторства тексту.

У такій моделі тексти є послідовністю символів, станів марковської ланцюга. Аналогічно в усному мовленні кожне слово можна описати за допомогою фонетичної транскрипції - послідовності фонем. Однак якщо при обробці текстів їхні стани (символи) відомі, то в звуковій мові спостерігаються не самі стани ланцюга (фонеми), а їх реалізації, тобто мовні сигнали, що представляють собою залежність звукового тиску від часу. Таким чином, самі стану-фонеми є прихованими: невідомо, яка фонема в дійсності була виголошена, відома тільки її реалізація. При цьому у зв'язку з варіативністю мови кожна фонема породжує безліч реалізацій. Таким чином, до традиційної для марковських ланцюгів задачі оцінки ймовірностей переходу від однієї

фонеми до іншої додається необхідність моделювання залежності спостережуваного сигналу від кожної конкретної фонеми.

У реальних системах розпізнавання мови замість фонем використовуються більш складні мінімальні звукові одиниці, такі як трифони - реалізації фонеми в контексті, кожна з яких описується за допомогою власної прихованої марковської моделі. Саме завдання побудови акустичної моделі - залежно акустичних характеристик реалізації мовних сигналів від типу звукової одиниці - є однією з найбільш складних при автоматичному розпізнаванні мови.

Приблизно до 2010 року на практиці використовувалася модель гауссових сумішей для завдання розподілу спостережуваного сигналу в залежності від фонеми. Для цього звуковий сигнал ділиться на невеликі ділянки (10-50 мс), для застосування традиційної обробки сигналів в частотній області для кожної ділянки сигналу виконується швидке перетворення Фур'є. Далі використовувалося логарифмування одержуваного спектра в зв'язку з відомим логарифмічним сприйняттям людським вухом, масштабу звуку. Нарешті, за допомогою дискретного косинусного перетворення логарифма спектра виходили практично незалежні ознаки - кепстральних коефіцієнти, розподіл яких і записувалося у вигляді суміші гауссовских випадкових векторів з діагональними ковариціонними матрицями.

Потім у зв'язку з революцією глибокого навчання замість традиційного підходу до вилучення характерних ознак і їх опису моделлю гауссових сумішей для побудови акустичної моделі мови стали використовувати глибокі нейронні мережі. У задачі розпізнавання мови застосовувалися звичайні мережі прямого поширення з великим числом шарів, які навчалися в режимі без вчителя послідовно від одного шару до іншого шару. Виявилося, що застосування такого підходу спільно з апаратом прихованих марковських моделей, що включають ймовірності переходу від однієї фонеми до іншої, на десятки відсотків підвищують точність розпізнавання спонтанної мови. Саме

цей підхід в даний час реалізований в більшості сучасних програмних бібліотек розпізнавання мови.

Поряд з появою нової акустичної моделі мови іншим проривним моментом стали нові мовні, лінгвістичні, моделі. У них в найпростішому випадку потрібно передбачити наступне слово по відомим попереднім словами - завдання, типове для обробки текстів. У традиційних системах застосовувалися моделі типу N-грам, в яких на основі великої кількості текстів оцінювалися розподілу ймовірності появи слова в залежності від N попередніх слів. Для отримання надійних оцінок розподілів параметр N повинен бути досить малий: одне, два або три слова - моделі уніграм, біграм або триграм відповідно.

Поява технологій глибокого навчання і розвиток рекурентних нейронних мереж для обробки текстів дозволили істотно поліпшити якість лінгвістичної моделі за рахунок обліку контексту і відсутність обмежень на використання тільки N попередніх слів. В результаті вийшло ще більше підвищити точність підсумкового розпізнавання мови - на слух можуть розпізнаватися не всі слова, і пропущені елементи важливо вгадувати по контексту, як це робить людина. Лінгвістичні моделі на основі рекурентних нейронних мереж, які дозволяють ефективно реалізувати таку поведінку, зараз повсюдно застосовуються в індустрії.

Аналізуючи ринок пристроїв неодноразово зустрічаються рішення які використовуються в Smart будинках, але вони мали не високу точність розпізнавання промовлених команд. Досить популярною є реалізація подібних систем на платформах Arduino, з використанням датчиків прийняття звуку. В представленому ж приладі реалізується сприйняття звуку з двох джерел, таких як модуль розпізнавання голосу та сприйняття звуку через мобільний термінал, що включає керування віддалено.

В питаннях покращення точності розпізнавання голосу в пристрою є рішення використання глибинних нейронних мереж (ГНМ), які в останні роки неодноразово показували суттєві результати процесах прогнозування,

класифікації, розпізнавання образів, рукописного тексту та мовлення. Тому використання ГНМ та їх модифікації у задачах розпізнавання мовлення – є актуальною задачею сьогодення.[11]

Існуючі підходи до розпізнавання мови: HMM

Попередні 30 років лідером у задачах мовлення вважались моделі, побудовані на основі прихованих ланцюгів Маркова(HMM)та Gaussian mixture model (GMM).[1] (рис.1.9)

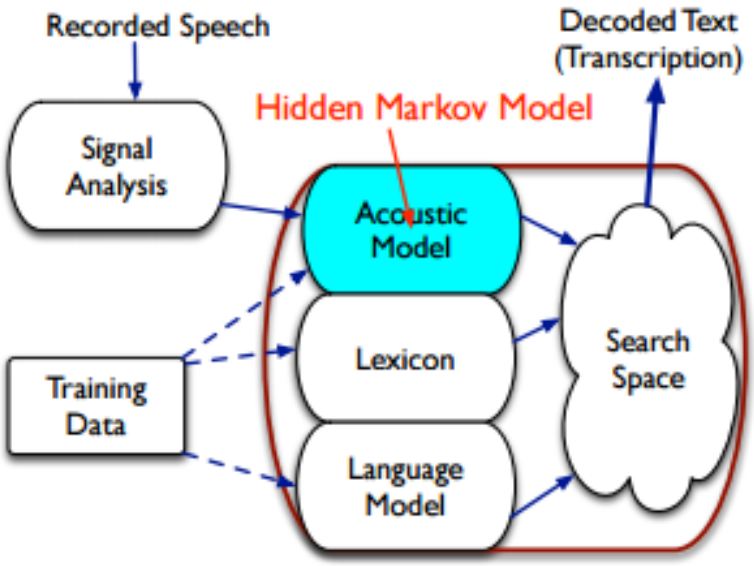


Рис.1.9 Схема використання HMM в розпізнаванні голосу

Записаний звук ділиться на короткі (10 мс) фрагменти, які аналізуються на вміст частот. Отриманий в результаті вектор характеристик пропускається через акустичну модель, яка видає набір імовірнісних розподілів серед всіх можливих фонем. HMM допомагає виявити послідовні структури в цьому наборі розподілів ймовірностей (рис.1.10).

Основний принцип тут повинен характеризувати слова в ймовірнісні моделі, де фонеми сприяють слову і являють стан HMM (Hidden Markov Modeling), в той час як ймовірності переходу були б ймовірністю наступної

виголошено фонеми, де моделі для слів, які є частиною словника, створюються в навчальній фазі.

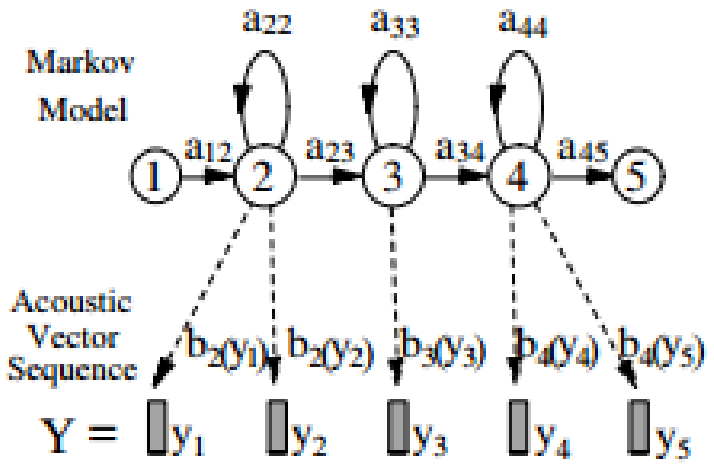


Рис.1.10 Виявлення послідовних структур за допомогою HMM

До переваг моделей на основі HMM відносяться:

•аналітичне вирішення проблеми розпізнавання;

•можливість розпізнавати слова, що складаються з набору букв без конкретного смислового значення;

•прості в реалізації та в навчанні.

Недоліки моделей на основі HMM:

•досить низька точність;

•погана робота в умовах шумує.

Існуючі підходи до розпізнавання мови: RNN (рис.1.11).

В останні 3-4 роки широкого використання набувають методи на основі рекурентних нейронних мереж (RNN).

Для побудови необхідної акустичної моделі з метою виділення фонем використовуються комірки довгої коротко-тривалої пам’яті (LSTM) у складі стандартних RNN.

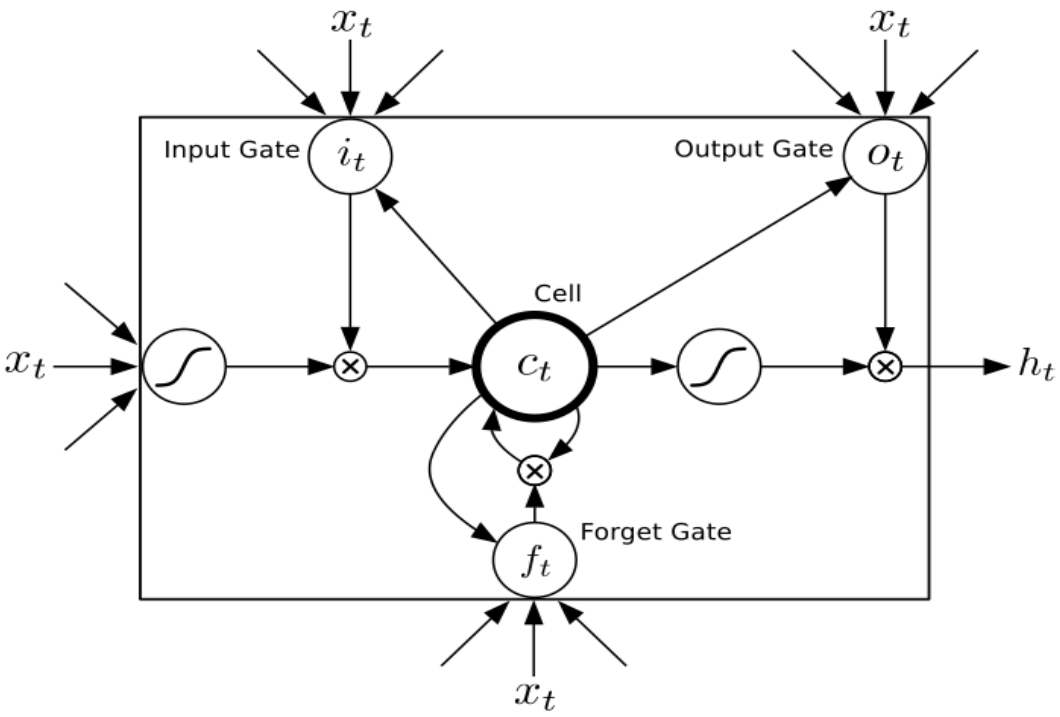


Рис.1.11 Існуючі підходи до розпізнавання мови RNN

Переваги моделей на основі RNN:

•більша швидкість роботи;

•точність розпізнавання більше;

•краще працюють в умовах підвищеного шуму;

•добре працює в умовах неточності та незавершеності промовлених слів.

Недоліки моделей на основі RNN.

•вимагає великих обчислювальні потужності;

•необхідна велика кількість прикладів для навчання;

•багато часу для навчання.

Висновки за розділом 1

Проведено аналіз існуючих методів розпізнавання мови людини. Були розглянуті основні типи задач, які не можуть існувати без систем розпізнавання. Теоретично обґрунтовано моделі і методи аналізу та розпізнавання сигналів багатьох змінних. Запропоновано методи, алгоритми і обчислювальні процедури аналізу сигналів на основі параметричних функцій систем, які створюють сигнал.

На сьогоднішній день існує багато методів вирішення цих проблем, але ні один метод не є ідеальним, їхня точність не перевищує [85%]. Метою є отримання результатів, які максимально будуть наближені до ідеальних.

РОЗДІЛ ІІ. РОЗРОБКА АЛГОРИТМУ РОЗПІЗНАВАННЯ ГОЛОСУ ЗА ДОПОМОГОЮ НМ

2.1 Опис роботи НМ в задачах розпізнавання мови.

2.1.1 Отримання даних з звукових сигналів

Порівнявши аналіз існуючих методів розпізнавання мови за допомогою нейронної мережі (НМ) було виявлено, що найпростішим методом навчання є завантаження аудіозаписів з яких буде проводитись аналіз (рис.2.1).

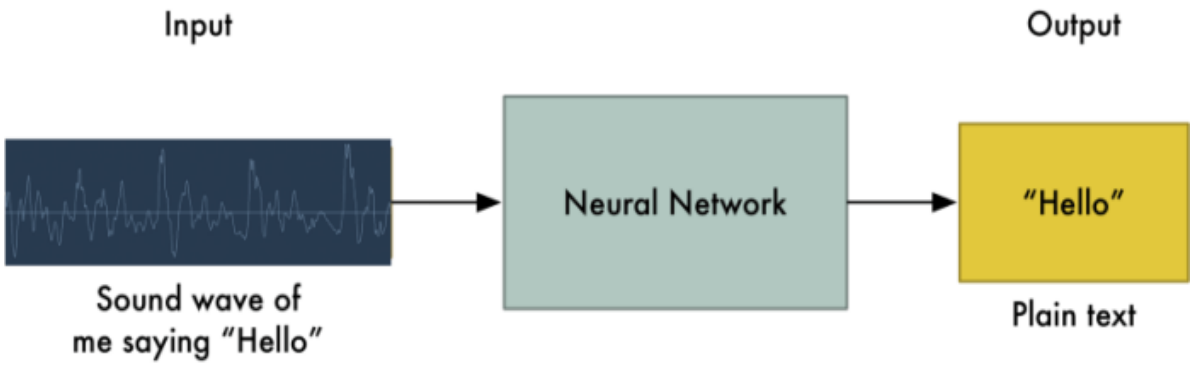


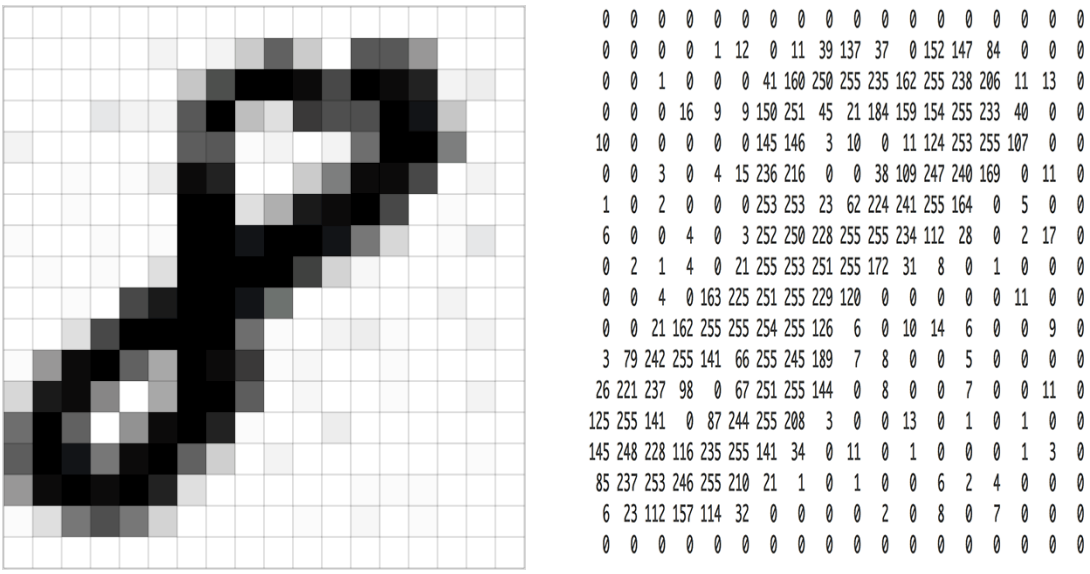
Рис.2.1 Проста схема розпізнавання мови за допомогою НМ.

При використанні такого методу виникає серйозна проблема, яка характеризується швидкістю мовлення диктора. Для прикладу одна людина вимовляє слово «привіт!» дуже швидко, а інша «пррррриииииввввііїт» повільно, при цьому створюючи довший звуковий файл з набагато більшим об’ємом даних. Із-за цього виникає складність з розпізнанням адже ці два файли мають бути розпізнанні як «привіт!». Автоматичне вирівнювання аудіофайлів різної довжини під фрагмент тексту фіксованої довжини - завдання непросте.

Щоб вирішити цю проблему, нам доведеться використовувати деякі спеціальні методи і додаткову точність в доповненні до глибокої нейронної мережі.

Перший крок в розпізнаванні мови - передача звуків на комп'ютер. Із матеріалів про НМ відомо, що для аналізу даних їх потрібно перевести в форму

зрозумілу для комп’ютера як в випадку з розпізнаванням числа 8 з картинки (рис.2.2).



0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	12	0	11	39	137	37	0	152	147	84	0	0	0
0	0	1	0	0	0	41	160	250	255	235	162	255	238	206	11	13	0
0	0	0	16	9	9	150	251	45	21	184	159	154	255	233	40	0	0
10	0	0	0	0	0	145	146	3	10	0	11	124	253	255	107	0	0
0	0	3	0	4	15	236	216	0	0	38	109	247	240	169	0	11	0
1	0	2	0	0	0	253	253	23	62	224	241	255	164	0	5	0	0
6	0	0	4	0	3	252	250	228	255	255	234	112	28	0	2	17	0
0	2	1	4	0	21	255	253	251	255	172	31	8	0	1	0	0	0
0	0	4	0	163	225	251	255	229	120	0	0	0	0	0	11	0	0
0	0	21	162	255	255	254	255	126	6	0	10	14	6	0	0	9	0
3	79	242	255	141	66	255	245	189	7	8	0	0	5	0	0	0	0
26	221	237	98	0	67	251	255	144	0	8	0	0	7	0	0	11	0
125	255	141	0	87	244	255	208	3	0	0	13	0	1	0	1	0	0
145	248	228	116	235	255	141	34	0	11	0	1	0	0	0	1	3	0
85	237	253	246	255	210	21	1	0	1	0	0	6	2	4	0	0	0
6	23	112	157	114	32	0	0	0	0	2	0	8	0	7	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Рис.2.2. Форма передачі даних для аналізу.

Так як звукові хвилі не так просто конвертувати в числові значення. Зробимо запис слова «Привіт»(рис.2.3).

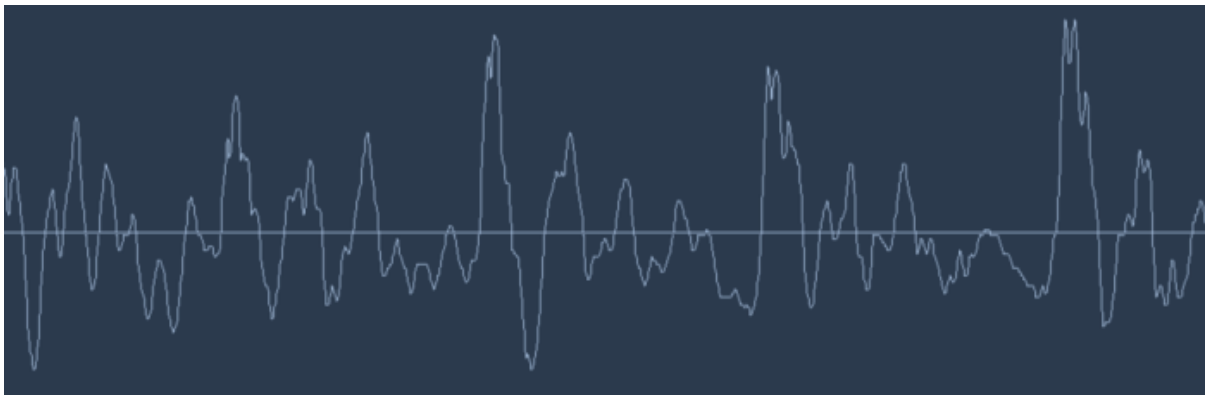


Рис.2.3. Вигляд звукового сигналу з словом «Привіт»

Звукові хвилі одномірні. У кожен момент часу у них є одне значення, залежне від амплітуди хвилі. Давайте наблизимо деяку невелику частину звукової хвилі і подивимося уважніше (рис.2.4).

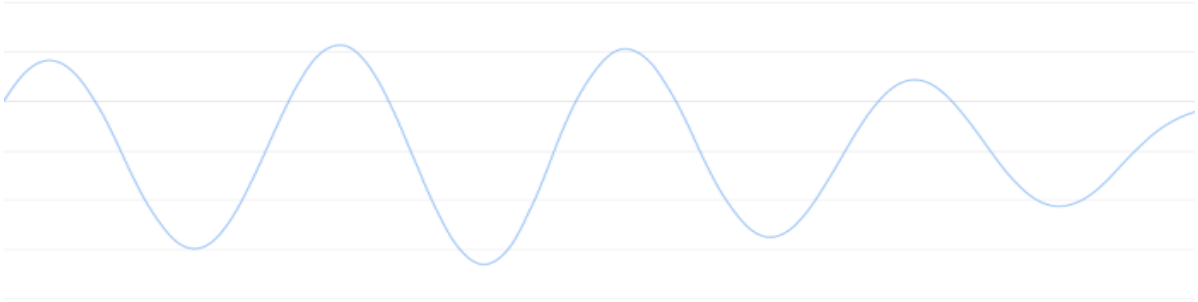


Рис.2.4. Збільшена частина звукової хвилі

Щоб перетворити цю звукову хвилю в числа, записуються значення амплітуди хвилі в рівновіддалених точках (рис.2.5, рис. 2.6.).

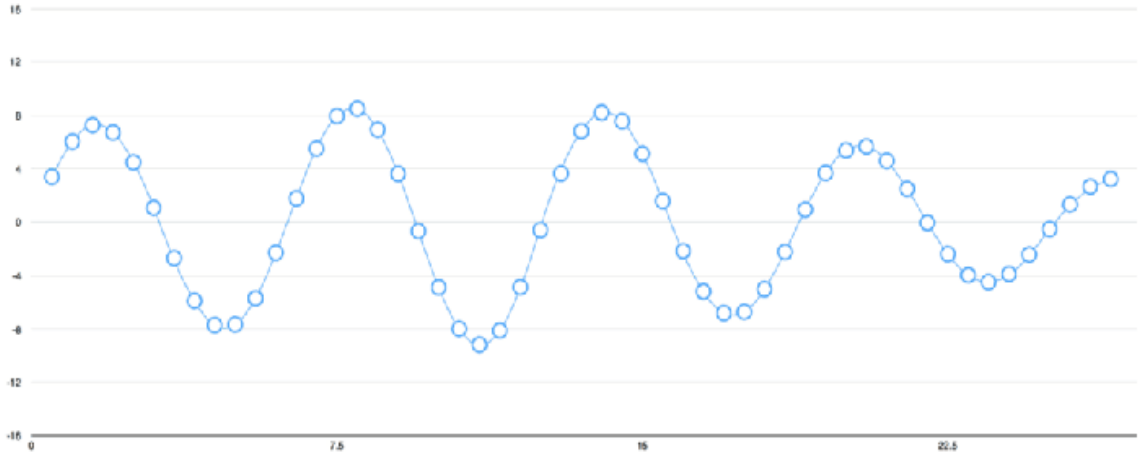


Рис.2.5. Розділ звукової хвилі на рівні частини.

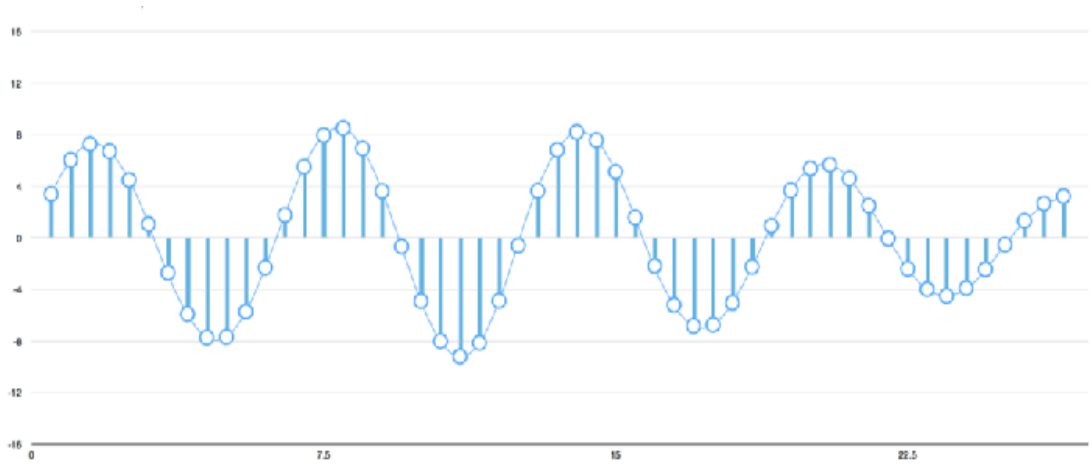


Рис.2.6. Замір значення хвилі в кожній з точок

Даний процес називається дискретизацією. Зчитування даних відбувається тисячі разів в секунду і записуємо числа, відповідні амплітуді звукової хвилі в цей момент часу. Виходять стиснені .wav аудіофайли.

Звук, що записується на CD-диски, має дискретизацію частотою 44,1кГц (44 100 відліків в секунду). Але для розпізнавання мови досить частоти дискретизації 16 кГц (16000 відліків в секунду), так як діапазон частот людської мови не настільки великий.

Тож візьмемо вихідний аудіозапис зі словом привіт і виконаємо його оцифрування з частотою 16 000 Гц/с. Отримаємо масив точок, ось перші 100 з них:

[-1274, -1252, -1160, -986, -692, -614, -286, -134, -57, -41, -169, -456, -450, -541, -761, -1067, -1231, -1047, -952, -645, -489, -448,

, -397, -212, 193, 114, -17, -110, 128, 261, 198, 390, 461, 772, 948, 1451, 1974, 2624, 3793, 4968, 5939, 6057, 6581, 7302, 7640, 7223, 6119,5461,

4820, 4353, 3611, 2740, 2004, 1349, 1178, 1885, 901, 301, -262, -499, -488, -707,- 1406, -1997, -2377, -2494, -2605, -2675, -2627, -2500, -2148,

-1648, -970, -364, 13, 260, 494, 788, 1011, 938, 717, 507, 323, 324, 325, 350, 103, -113, 64, 176, 93, -249, -461, -606, -909, -1159, -1307, -1544]

Іноді з’являються сумніви, що дискретизація створює лише наближений варіант вихідної звукової хвилі, так як зчитуються випадкові свідчення, а в проміжках між відліками дані втрачаються. Але завдяки теоремі Котельникова відомо, що для ідеального відтворення вихідної звукової хвилі досить використовувати частоту дискретизації, що вдвічі перевищує найвищу частоту записуваного звуку.

$$x(t) = \sum_{k=-\infty}^{\infty} x(k\ \Delta) sinc\left[\frac{\pi}{\Delta}(t - k\ \Delta)\right],$$

де $sinc(x) = \sin(x)/x$ - функція sinc. Інтервал дискретизації задовольняє обмеженням $0 < \Delta \leq 1/2f_c$. Миттєві значення даного ряду є дискретні відліки сигналу.

На цьому зроблено акцент тільки тому, що майже всі помилково думають, що використання більш високих частот дискретизації завжди призводить до кращої якості звуку. Це не так.

2.1.2 Обробка отриманих оцифрованих даних

Тепер наявний масив чисел, кожне з яких представляє амплітуду звукової хвилі через інтервали 1/16000 секунди.

Є можливість навчити НМ на цих числах, але розпізнання мовних моделей шляхом обробки цих чисел безпосередньо важко. Замість цього є можливість полегшити задачу, провівши попередню обробку звукової інформації.

Почнемо з того, що згрупуємо відліки під фрагменти по 20 мілісекунд. Ось перший такий фрагмент (перші 320 відліків):

[-1274, -1252, -1160, -986, -692, -614, -286, -134, -57, -41, -169, -456, -450, -541, -761, -1067, -1231, -1047, -952, -645, -489, -448,

, -397, -212, 193, 114, -17, -110, 128, 261, 198, 390, 461, 772, 948, 1451, 1974, 2624, 3793, 4968, 5939, 6057, 6581, 7302, 7640, 7223, 6119,5461,

4820, 4353, 3611, 2740, 2004, 1349, 1178, 1885, 901, 301, -262, -499, -488, -707,- 1406, -1997, -2377, -2494, -2605, -2675, -2627, -2500, -2148,

-1648, -970, -364, 13, 260, 494, 788, 1011, 938, 717, 507, 323, 324, 325, 350, 103, -113, 64, 176, 93, -249, -461, -606, -909, -1159, -1307, -1544

-1815, -1725, -1341, -971, -959, -723,-261, 51, 210, 142, 142, -92, -345, -439, -529, -710, -907, -887, -693, -403, -180, -14, -12, 29, 89, -47...]

Побудова цих чисел у вигляді простого лінійного графіка (рис.2.7) дає приблизне зображення вихідної звукової хвилі за обраний 20 мілісекундуного періоду часу:

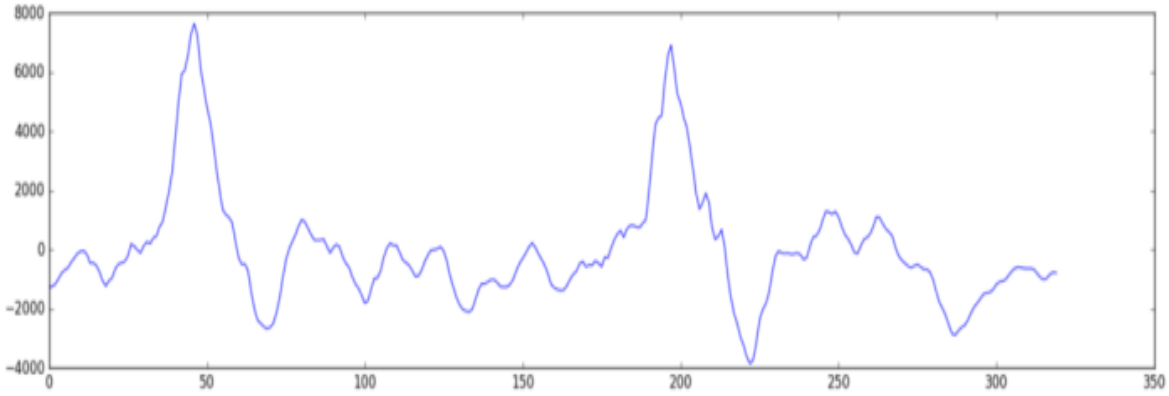


Рис.2.7. Графік вихідного оцифрованого голосового сигналу за 20 мсек.

Цей запис триває всього 1/50 секунди. Але навіть він являє собою складну суміш різних частот звуку. Є кілька низьких звуків, є середньо частотні звуки і навіть деякі високі звуки. Всі ці частоти змішуються разом - і виходить звук людської мови.

Щоб спростити обробку цих даних для нейронної мережі, розкладемо складну звукову хвилю на її складові частини, починаючи від найнижчих частот. Потім, підсумовуючи потужності звуку в кожній смузі частот, створюємо частотну картину звуку. Для прикладу, якщо взяти запис того, як хтось виконує акорд До-мажор на фортепіано. Цей звук являє собою комбінацію з трьох музичних нот - До, -Мі і -Соль - які змішуються в один складний звук. Треба розбити цей складний звук на окремі ноти, щоб виявити вихідні ноти. В записі - те ж саме.

Це робиться за допомогою математичної операції, перетворення Фур'є. Відбуваєтьсярозкладення складної звукової хвилі на прості звукові хвилі, які її складають. Маючи окремі звукові хвилі, складаються потужності звуку в кожній з них.

Кінцевим результатом є оцінка важливості кожного частотного діапазону, від низьких частот до високих. Числа нижче описують потужність звуку в кожній смузі по 50 Гц у вихідному фрагменті рис.2.8.:

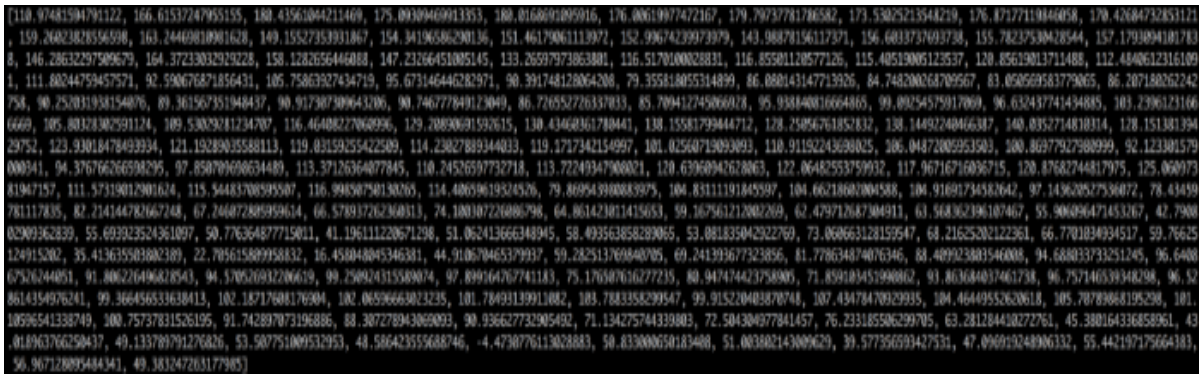


Рис.2.8 Масив даних з значеннями потужності звуку в смузі по 50 Гц.

Для того щоб зрозуміліше описати картину побудуємо діаграму (рис. 2.9) :

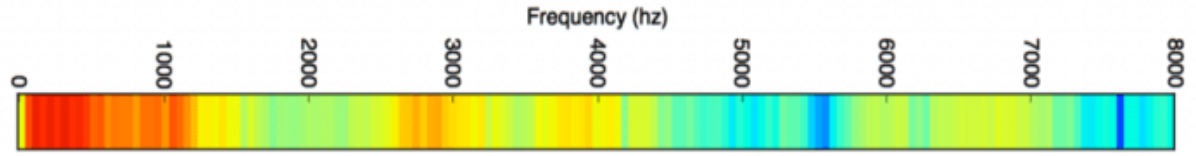


Рис.2.9 Діаграма по отриманих значеннях

Якщо цей процес повторюється на кожному 20-мілісекундному фрагменті аудіо, буде отримано спектрограму (кожен стовпець зліва направо є один 20-мілісекундний фрагмент) (рис. 2.10).

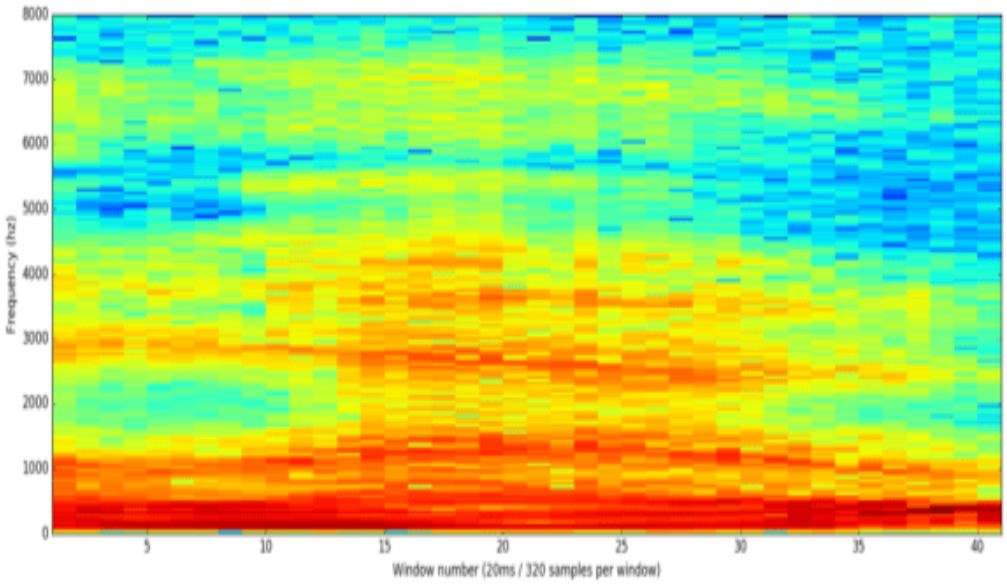


Рис.2.10. Спектограмма запису

За допомогою спектограми можна виділити музичні ноти й інші тони в аудіо. Нейроннії мережі буде простіше знаходити шаблони в таких даних, ніж в сирих записах звуку. Тепер сформованно дані, які передаються нейронній мережі.

2.1.3 Розпізнавання букв з коротких звуків.

Тепер, коли є аудіо в форматі, з яким можна працювати далі, можна навчати на цих даних глибоку НМ. НМ буде отримувати аудіо-фрагменти довжиною 20 мс. Для кожного невеликого фрагмента мережа спробує визначити, яка буква була виголошена рис.2.11.

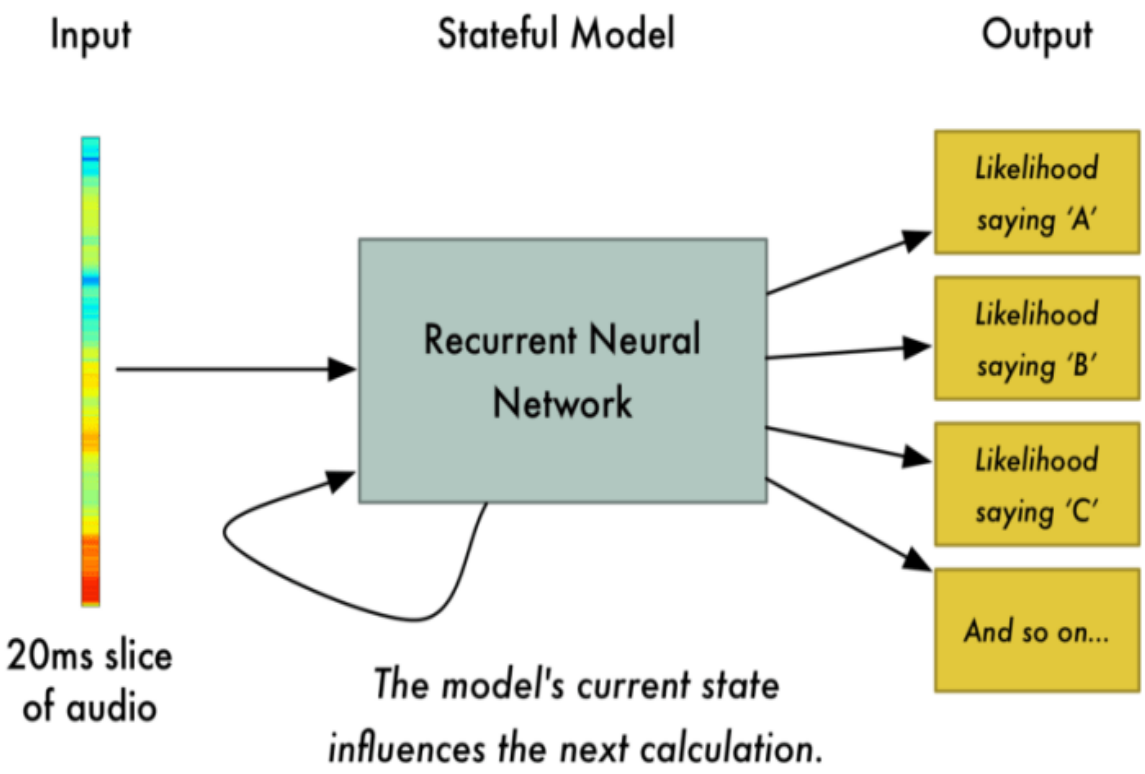


Рис.2.11 Схема обробки аудіозапису

Краще використовувати рекурентну нейронну мережу, тобто таку, яка на кожному кроці враховує результати попередніх кроків. Її перевагою є можливість впливати на ймовірну наступну букву, буквою попередньо визначеною мережею. Наприклад, якщо сказано «прив», то швидше за все,

далі слідує «іт», щоб закінчити слово «Привіт». Набагато менш ймовірно, що буде сказано щось невимовне, наприклад «ШХЩ». Таким чином, запам'ятовуючи попередні результати, НМ зможе робити більш точні прогнози в майбутньому.

Після того як нейронна мережа обробить весь аудіо файл (по одному фрагменту за раз), буде отримано розкладання кожного фрагмента аудіо на літери, найбільш ймовірно виголошені під час цього фрагмента. Відображення «Привіт» наведено на рис. 2.12.

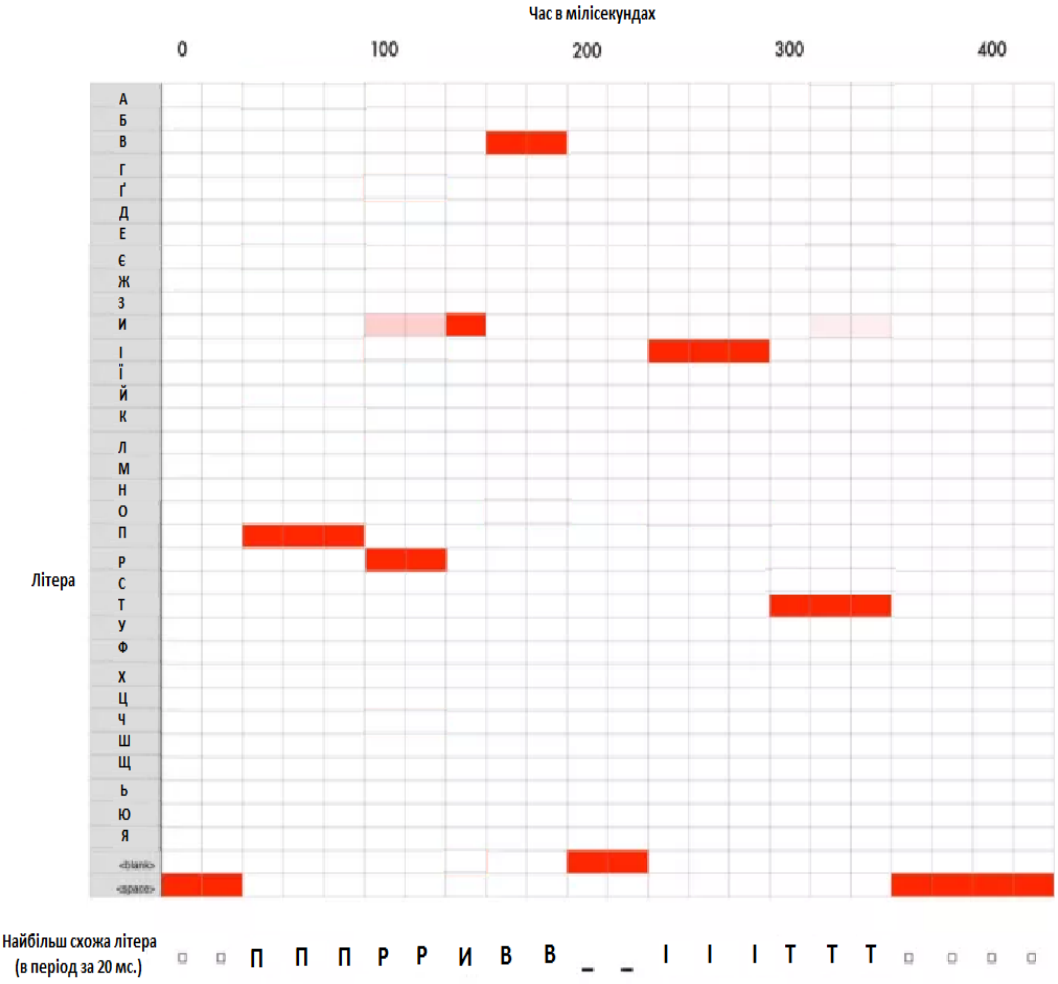


Рис 2.12 Розкладання промовленого слова на літери

Нейронна мережа прогнозує що було вимовлено «ПППРРИВВ__ІІТТТ». Але також вона може подумати, що було сказано «ПРРРИИВВВ__ВВ_ІІД» або навіть «ПРІУУ_УУУІІІТ».

Результати будуть систематизовані у кілька етапів. По-перше, заміна будь-яких повторюваних символів одним символом:

- ППППРРИВВ__ІІТТТ перетвориться в ПРИВ_ІТ
- ПРРРИИВВВ__ВВ_ІІД перетвориться в ПРИ_В_ІД
- ПРІУУ_УУУІІІТ перетвориться в ПРІ_У_ІТ

По-друге, видалення всіх пропусків:

- ПРИВ_ІТ перетвориться в ПРИВІТ
- ПРИ_В_ІД перетвориться в ПРИВІД
- ПРІ_У_ІТ перетвориться в ПРІУІТ

Отже, отримано три можливих транскрипції - «Привіт», «Привід» і «Пріуіт». Якщо їх вимовити, всі вони будуть звучати схоже на «Привіт». Оскільки всі букви визначаються по одній, НМ може вигадати абсолютно невимовні транскрипції. Наприклад, якщо вимовити «Він не піде», мережа може видати «Уін небі те».

Принцип розпізнавання за допомогою НМ полягає в тому, щоб уточнювати ці передбачення, порівнюючи їх з великою базою даних письмового тексту (книги, новинні статті і т. п.). Найменш вірогідні транскрипції відкидаються і надається перевага тим, що здаються найбільш реалістичною.

З трьох варіантів транскрипцій «Привіт», «Привід» і «Пріуіт», очевидно, «Привіт» зустрічається в базі даних тексту більш часто (не говорячи вже про вихідних аудіо файл), і тому, ймовірно, це правильний варіант. Тому обираємо «Привіт» в якості остаточної транскрипції.

Разом з цим виникає інша складність якщо хтось скаже «Привід»? Це буде вірне слово, і «Привіт» буде неправильною транскрипцією! Звичайно, можливо, що хтось насправді скаже «Привід» замість «Привіт». Але система розпізнавання мови, подібна до цієї (навчається на літературній мові), тому не вибере «Привід» як правильний варіант. Крім того, коли ви говорите «Привід», ви все одно маєте на увазі «Привіт», навіть якщо ви підкреслюєте букву «Д» . Якщо ваш телефон налаштований на розпізнавання мови,

спробуйте сказати телефону «Привід». Він відмовиться розуміти вас, і буде розпізнавати це як «Привіт».

Google використовує в якості бази даних всі запити які були здійснені через його сервіси, якщо ви користувались голосовим керуванням то ви можете перейди по посиланню і побачити всі запити на Okay Google які були здійснені з вашого телефона (https://myactivity.google.com/udc/vaa)[4].

2.2. Розробка алгоритму для аналізу голосу

2.2.1. Короткий опис розробки алгоритму рішення

Розглянемо тут рішення практичного завдання управління приладами та отримання даних за допомогою голосових команд. Причому вибір команд - окремих слів виконаний так, щоб слова були близькі за вираженням. Це необхідно для оцінки стійкості представленого рішення розпізнавання команд. Введення команд здійснюється через стандартний мікрофон, підключений через популярний аудіо адаптер CMI 8738 / PCI до комп'ютера, що працює під керуванням операційної системи Linux Ubuntu 10.04. Рішення завдання розбивається на етапи (рис.2.13.):

1. Виділення із загальної звукової оцифрованої осцилограми тривалістю 2 секунди осцилограми конкретного слова - команди.

2. Розбиття осцилограми на окремі ділянки довжиною ~ 15-23 мілісекунд (довжина команди - слова ~ 0.65-0.90 секунди).

3. Застосування дискретного перетворення Фур'є (ДПФ) [7] до кожної ділянки слова (отримання спектра сигналу на ділянці).

4. Виділення на кожній ділянці n - точок локальних максимумів амплітуд з їх значеннями частот (виділення формант мовного сигналу).

5. Пошук на основі зарані отриманих результатів такого значення n, при якому після відновлення звуку за допомогою синусового перетворення Фур'є буде однозначне суб'єктивне "впізнавання" перетвореного слова з набору вибраних команд.

6. Формування масиву даних для кожного слова з ділянок слова, які будуть характеризувати конкретне слово.

7. Так як одне і те ж слово, повторене навіть одним і тим же диктором, має осцилограми, що відрізняються один від одного, то створюється набір з декількох десятків масивів, характерних для одного і того ж слова.

8. Отримавши окремі набори масивів характерних для конкретних слів, використовується математичний апарат нейронних мереж для розпізнавання конкретного введеного через мікрофон слова. Тут для створення і роботи з нейронною мережею використовується широко поширена бібліотека FANN [2].

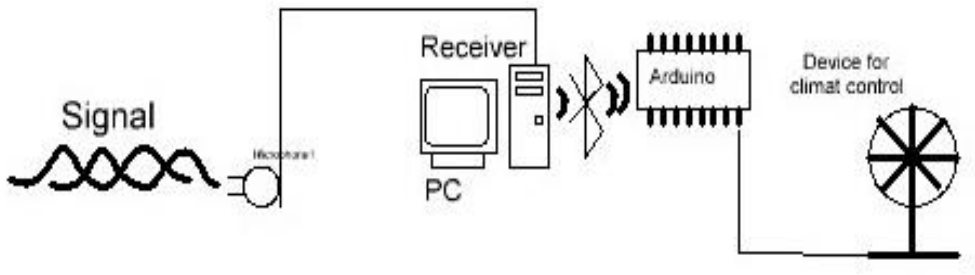


Рис 2.13 Схема пристроїв для проведення досліджень

Розглянемо рішення задачі в представленій вище послідовності.

1. Виділення з осцилограми конкретного слова - команди.

Для запису голосу можна скористатися відомою в Linux Ubuntu командою arecord. За допомогою командного рядка:

arecord -q -d 2 -f cd -r 16000 -c 1 a.wav

- arecord – виклик команди з ПО Ubuntu.
- q – аргумент для шумопригнічення.
- d – аргумент котрий визначає тривалість запису.
    - f cd – аргумент котрий ставить формат квантування
    - r (rate) – частота дискретизації
    - c – канал запису, по дефолту 1.

- a – назва файла.
- .waw – формат запису

створюється 2-х секундний монофонічний файл a.wav з частотою дискретизації 16000Гц і розрядністю або квантуванням (-f cd) в 16 біт.

На рис.3.1 представлена осцилограма слова "зелений", відображена в аудіо редакторі audacity [3]. Тут по осі абсцис представлено час в секундах, по осі ординат - нормоване значення амплітуди.

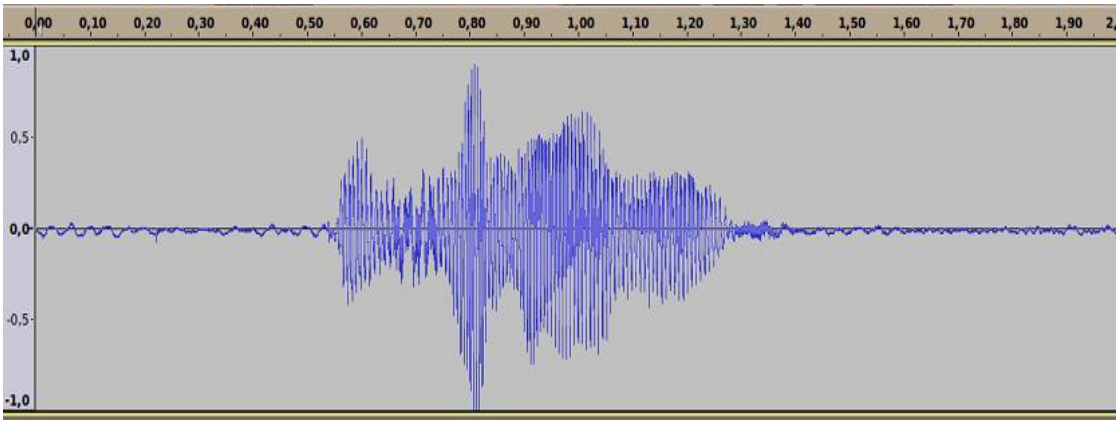


Рис.2.14. Осцилограма слова "зелений" в 2-х секундному аудіо файлі a.wav

Для виділення осцилограми слова "зелений" складена програма на мові c++ slice.c, текст якої представлений в Додатку А. Працює програма наступним чином (рис.1.15):

- Зчитується файл a.wav.

- Визначається максимальна і мінімальна амплітуда осцилограми (для розрядності 16 біт максимальні і мінімальні значення не можуть перевищувати +32 767 і - 32 768 відповідно).

- Виконується нормалізація амплітуд (розподіл всіх позитивних амплітуд на максимальну амплітуду, негативних - на мінімальну).

- осциллограмма розбивається на 200 ділянок.

- Для кожної ділянки визначається середнє позитивне значення амплітуди.

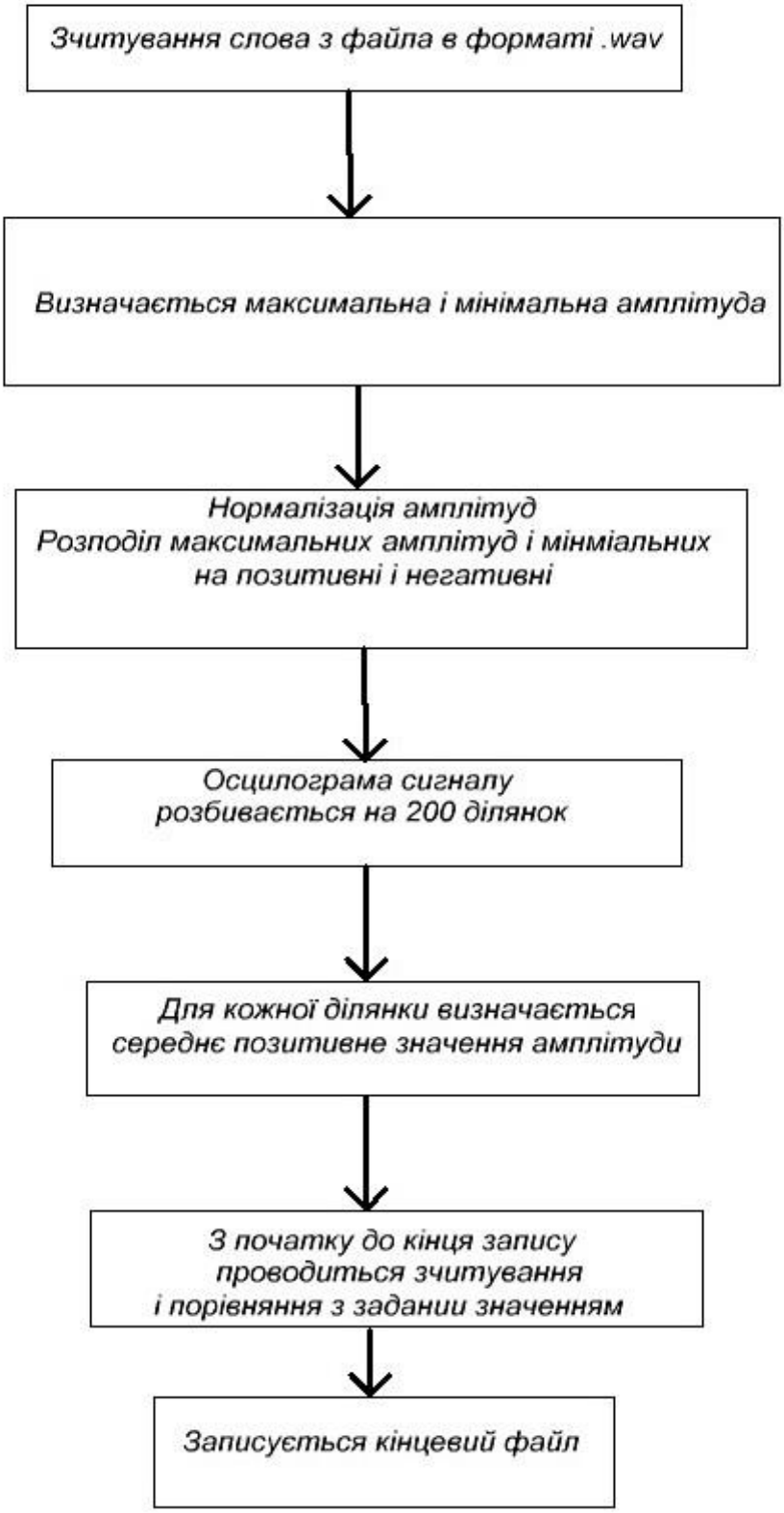


Рис.2.15 Блок-схема роботи програми slice.c

- Виконується "прогін" по ділянках осцилограми ліворуч і праворуч. Якщо значення середньої позитивної амплітуди більше заданої величини (5% з урахуванням зовнішніх шумів), то передбачається, що початок і кінець слова досягнуто.

- Запис виділеного слова "зелений" в файл c.wav.

На рис.2.16 представлена осцилограма виділеного слова "зелений". Видно, що його тимчасова довжина приблизно дорівнює 0.73с.

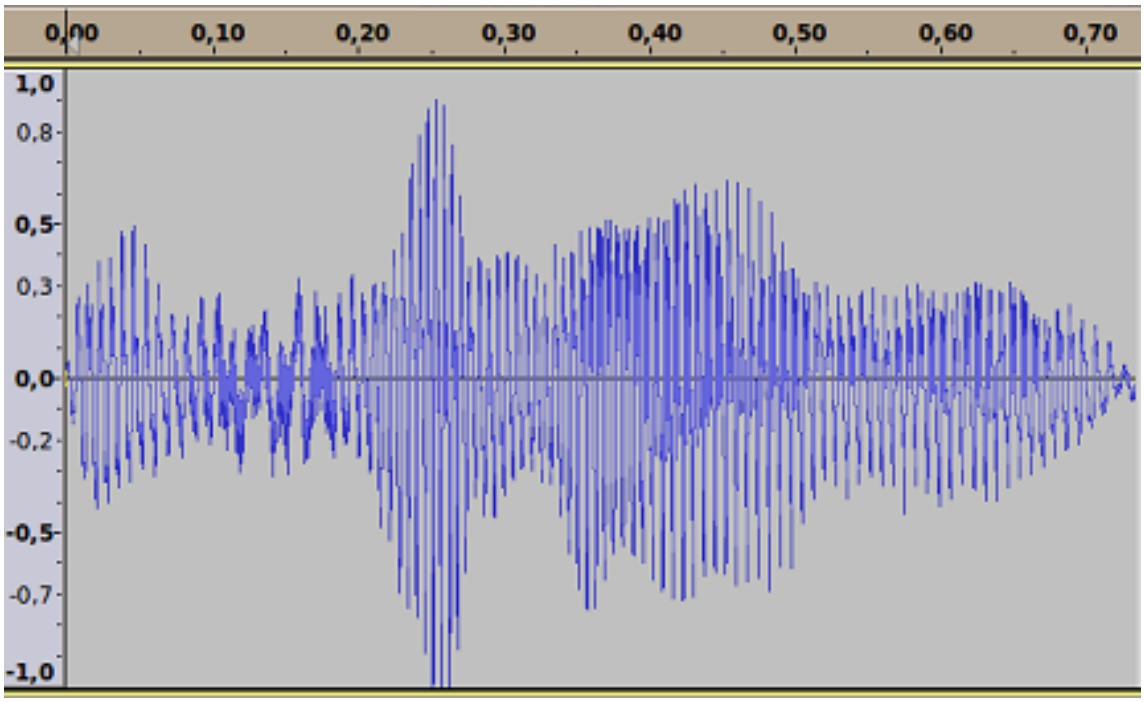


Рис.2.16. Осцилограма слова "зелений" виділеного програмою slice.c

2.2.2. Спектральний аналіз сигналу

Для розпізнавання слова необхідно знайти масив чисел, який би представляв собою визначене слово. У роботі це зроблено в такий спосіб. Ділимо одне з заданих слів на 40 інтервалів. У цьому випадку довжина інтервалу знаходиться в межах 15 ... 23МС. До кожного інтервалу застосовуємо дискретне перетворення Фур'є.

Відомо, що пряме перетворення Фур'є записується у вигляді [1]:

$$X_k = \frac{1}{N}\sum_{n=0}^{N-1} x_n e^{-\frac{2\pi ikn}{N}}$$

$$k = 0, \dots, \frac{N-1}{2}$$

де: N - кількість значень сигналу, виміряних в одному з 40-ка інтервалів;

x_n - виміряні значення сигналу в n - ій точці інтервалу;

X_k - комплексна амплітуда;

k-й синусоїдальний сигнал (k - й індекс частоти на кривій спектру).

Індекс k змінюється від 0 до $\frac{N-1}{2}$ так як друга половина з N комплексних амплітуд, фактично, є дзеркальним відображенням першої і не несе додаткової інформації.

Розкладемо експоненту по формулі Ейлера і отримаємо:

$$X_k = \frac{1}{N}\left[\sum_{n-0}^{N-1} x_n \cos\left(\frac{2\pi kn}{N}\right) - i\sum_{n-0}^{N-1} x_n \sin\left(\frac{2\pi kn}{N}\right)\right];$$

або:

$$X_n = \frac{1}{N}(ReX_k + ImX_k).$$

Визначення дійсності (речової) амплітуди виконується за формулою:

$$A_k = (\frac{1}{N})\sqrt{ReX_k^2 + ImX_k^2}$$

частоти:

$$Fr_k = (\frac{1}{T})$$

T - період часу на протязі якого обрано вхідні дані (тривалість одного з 40-ка інтервалів).

Геометрична інтерпретація представлена на рис.2.17.

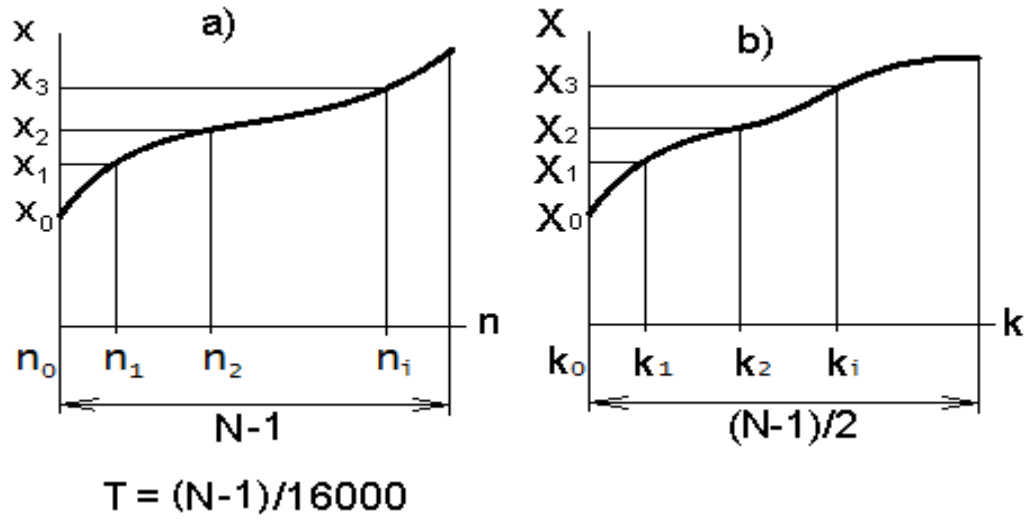


Рис.2.17. Геометрична інтерпретація дискретного перетворення Фур'є.
а) - осцилограма інтервалу (n - індекс часу), б) - спектрограмма (k - індекс частоти).

Фрагмент програми ДПФ має вигляд (програма new.c [Додаток Б.]):

…

```
N=yy/nom; // yy – число точок дискретизації команди, nom=40 – число
// інтервалів, N – кількість точок в інтервалі
tt0=nn*N; // nn – номер інтервала (nn = 0, …, 39)
tt1=(nn+1)*N; // tt0, tt1 – номера точок початку і кінця інтервалу
Tf=float(tt1-tt0)/16000; // Тривалість інтервалу в с.
  {
j=0;
      Nf=tt1-tt0; // Кількість вимірювань (так само N)
   for (kf=0; kf<Nf/2; kf++) // kf – відповідає k (індекс частоти)
   {
     summa_Re=0;  summa_Im=0;
         for (nf=0; nf<Nf; nf++) // nf – відповідає n (індекс часу)
         {
          Arg=2*3.141592653589793*nf*kf/Nf;
          Re=cos(Arg)*an1[nf]/Nf;
             Im=sin(Arg)*an1[nf]/Nf;
       summa_Re=summa_Re+Re; // Расчет ReXk
```

```
            summa_Im=summa_Im+Im; // Расчет  ImXk
        }
 // Розрахунок дійсної амплітуди Ak і частоти Frk
        A[kf]=1.5*(sqrt((summa_Re)*(summa_Re)+(summa_Im)*(summa_Im)));
        Fr[kf]=(1/Tf)*kf;
  }
…
```

Як приклад на рис.2.18., представлена осцилограма команди «температура», відкрита в програмі audacity. Затемнений вертикальний стовпець відповідає інтервалу №12.

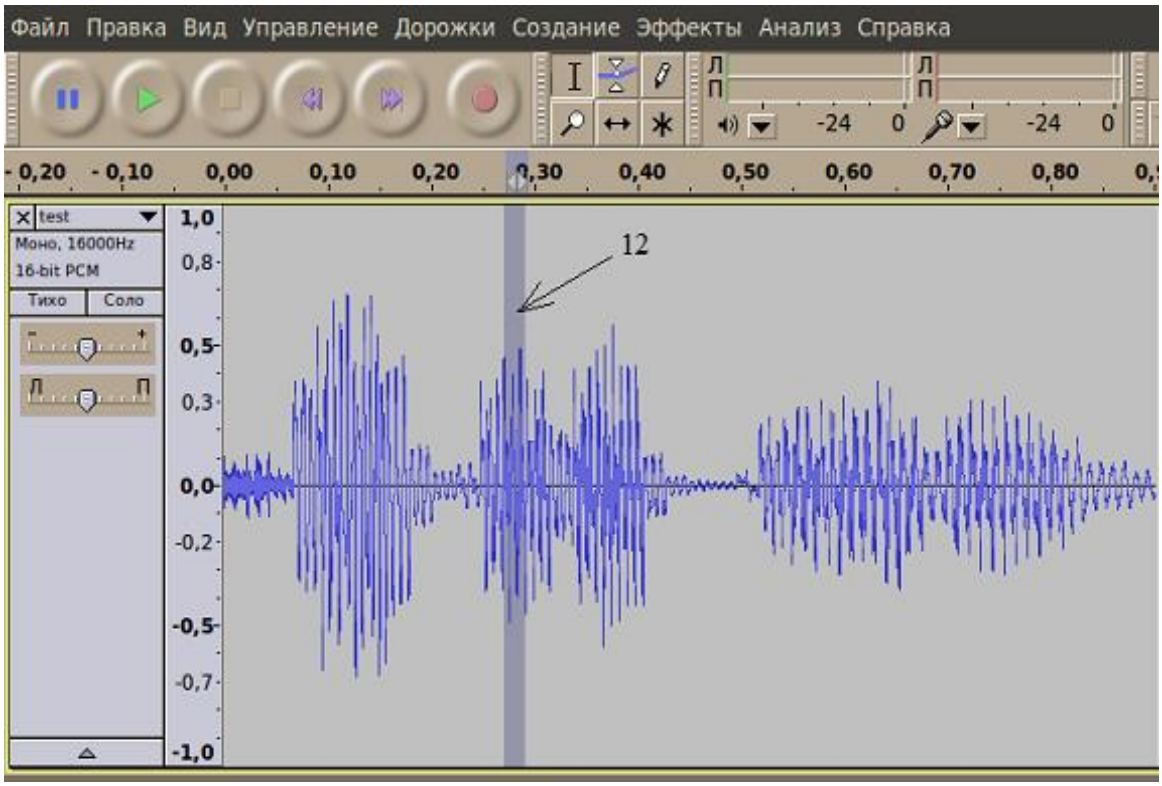


Рис.2.18. Осцилограма слова «температура» з виділеним інтервалом №12

За допомогою програми new.c розрахована спектрограмма для цього інтервалу і для діапазону частот 0, ..., 2500Гц, яка показана на рис.2.19. Амплітуди представлені цілими числами 16-ти бітної розрядності, як вони представлені в .wav файлі.

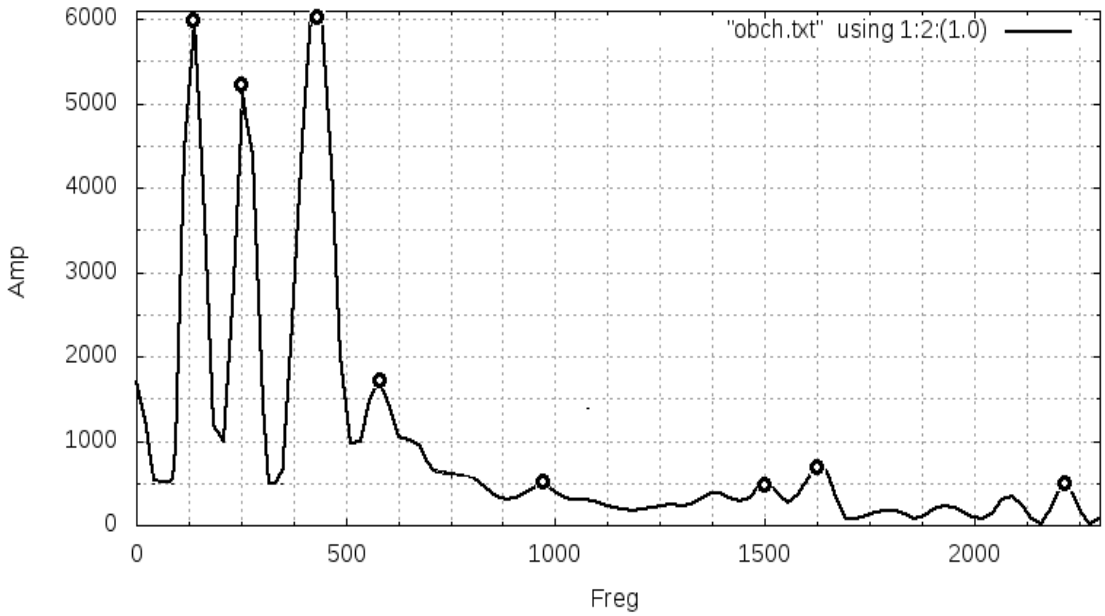


Рис.2.19. Спектрограма інтервалу №12 слова «температура»

На спектрограмі кружечками відзначено 8-ім локальних екстремумів (формантів), що мають найбільше значення в цьому інтервалі. У роботі передбачається, що для опису слова в кожному з 40-а інтервалів досить обчислити значення 8-ми векторів (кожен задається парою значень - частотою і амплітудою). Таким чином, слово однозначно для невеликої бази словникового запасу описується масивом з 640 чисел. Доказ цього виконувався експериментально шляхом відновлення осцилограми слова - команди з отриманої спектрограми функціями синуса по 8-ми амплітудам і частотам. Програма new.c формувала файл test.wav, який згодом прослуховувався. Якщо при програванні файлу звучання суб'єктивно відповідало вимовленому слову, то робився висновок про можливість опису слова таким набором з 8-ми векторів. Зменшення кількості векторів призводить до зменшення суб'єктивної розрізнення слова.

Нижче представлений фрагмент програми відновлення звуку з 8-ми формантів за допомогою функцій синуса.

…

```
float pi=2.*3.14159265; float di;
 for (i=tt0; i<tt1; i++) // tt0, tt1 – номера точок початку і кінця інтервалу
```

```
    {    di=pi*i/16000;
// buffer[i] – значення амплітуд поля даних wav - файлу
// max0, …, max7 – максимальні амплітуди в інтервалі спектра
// maxf0, …, maxf7 – частоти максимальних амплітуд в інтервалі спектра
buffer[i] =(int)(max0*sin(di*maxf0))+(int)(max1* sin(di*maxf1));
buffer[i]=buffer[i]+(int)(max2 * sin(di*maxf2))+(int)(max3 * sin(di*maxf3));
buffer[i]=buffer[i]+(int)(max4 * sin(di*maxf4))+(int)(max5 * sin(di*maxf5));
buffer[i]=buffer[i]+(int)(max6 * sin(di*maxf6))+(int)(max7 * sin(di*maxf7));
    }
…
bbuf=tt1; // кількість елементів масиву buffer
write_wav1(namefil, bbuf , buffer, 16000); // Запис wav файла
...
```

Результатом роботи цього фрагмента програми new.c є формування і запис wav файлу з ім'ям test.wav. На рис.2.20 представлена осцилограма звуку після синусоидального відновлення слова «зелений». Зіставляючи осцилограми на рис.2.18., і рис.2.20., можна помітити їх близьку подібність.

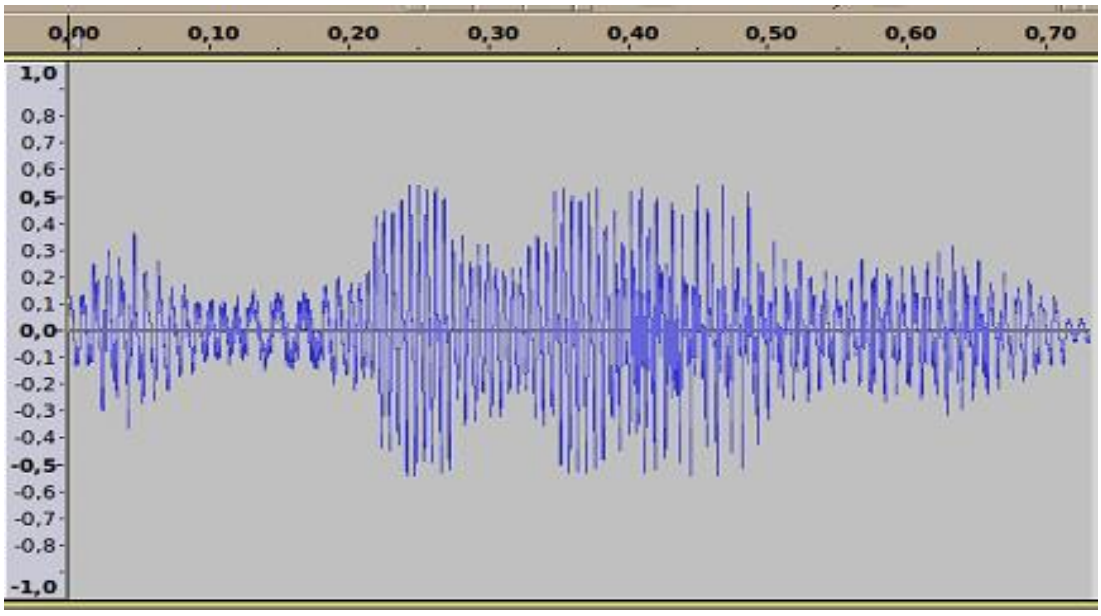


Рис.2.20. Осцилограма звуку після синусоідального відновлення слова «зелений»

Результатом роботи програми new.c є також формування масиву з 640 чисел (320 векторів), що визначають вимовлене слово і запис його в файл r1.txt. Нижче представлений фрагмент програми:

```
        // Цикл по інтервалах
        i6=0;
for(nn=0;nn<nom;nn++)
{
…  amm[i6]=max0;amm[i6+1]=max1;amm[i6+2]=max2;
amm[i6+3]=max3;amm[i6+4]=max4;amm[i6+5]=max5;
amm[i6+6]=max6;amm[i6+7]=max7;
fmm[i6]=maxf0;fmm[i6+1]=maxf1;fmm[i6+2]=maxf2;
fmm[i6+3]=maxf3;fmm[i6+4]=maxf4;fmm[i6+5]=maxf5;
fmm[i6+6]=maxf6;fmm[i6+7]=maxf7;
i6=i6+8;
i7=i6; …
}
// Кінець циклу по інтервалах
// Формування масиву амплітуд і частот для сказаного слова і
// запис його в файл r1.txt
…
 FILE *filem;
filem = fopen("r1.txt","w");
for (i=0; i<i7; i++)
fprintf(filem,"%0.7f   %0.7f\n",amm[i],fmm[i]);
fclose(filem);
```

Для того щоб показати ефективність такого уявлення слів, слова-команди вибиралися близькі за вимовою:

- «синій», «сильний» - виконують команди включити синій і вимкнути синій;

- «червоний», «класний» - включити червоне і вимкнути червоний;

- «зелений», «земля» - включити зелене і вимкнути зелений;

- «температура» - отримати значення температури з датчика температури.

За цими командами контролер Arduino [5] повинен включати і вимикати три пристрої, отримувати і передавати на комп'ютер дані з температурного датчика. Комп'ютер і Arduino пов'язані з допомогою Bluetooth пристроїв. При відтворенні вихідних звукових файлів - команд і їх звуки після ДПФ і відновлення синусом. Якщо прослухати отримані звуки, можна виявити їх відмінність один від одного і суб'єктивно встановити відповідність їх вимовленим командам. Отже, можна виконати кілька десятків повторень однієї і тієї ж команди і сформувати для неї кілька десятків масивів (r1.txt). Те ж зробити для всіх інших команд. Тоді буде отримано наборів даних, кожен з яких відповідає своєму слову-команді. За отриманими наборам даних можна виконати навчання нейронної мережі. З появою нового набору даних (проголошенням команди) за допомогою нейронної мережі можна встановити відповідність, до якого набору даних відноситься новий набір. Таким чином виконати розпізнавання слова.

2.2.3. Створення за допомогою бібліотеки FANN нейронної мережі для розпізнавання команд

Виконаємо установку бібліотеки FANN для Linux Ubuntu 10.04. Робота з нейронною мережею буде виконуватися в скриптовій мові PHP.

Встановлюємо пакети:

sudo apt-get install php5-cli

sudo apt-get install php5-dev

Далі встановлюємо бібліотеки FANN 1-й версії:

sudo apt-get install libfann1

sudo apt-get install libfann1-dev

Копіюємо з сайту http://pecl.php.net/package/fann Wrapper for FANN (для мови PHP) - програму fann-0.1.1.tgz

Розпаковуємо її:

gzip -d fann-0.1.1.tgz

tar xf fann-0.1.1.tar

cd fann-0.1.1 /

запускаємо команди

phpize

./configure

Виконуємо компіляцію:

make

Виправляємо помилки компіляції. Для цього редагуємо файл php_fann.h - необхідно закоментувати рядок 28

#define PHP_FANN_OO 1

Далі компілюємо заново

make

і додаємо в php.ini рядок

extension = fann.so

Після цього скрипт, написаний на PHP буде працювати з бібліотекою FANN.

Дані з дослідження показали, що в якості мережі можливе використання класичної взаємопов’язаної багатошарової штучної нейронної мережі, заснованої на алгоритмі навчання зі зворотним поширенням помилок.

Навчання мережі методом зворотного поширення помилки включає в себе три етапи: подачу на вхід даних, з подальшим поширенням даних в напрямку виходів, обчислення і зворотне поширення відповідної помилки для коригування ваг. Після навчання передбачається лише подача на вхід мережі даних та поширення їх в напрямку виходів. При цьому, якщо навчання мережі може бути досить тривалим процесом, то безпосереднє обчислення результатів навченої мережі відбувається дуже швидко. Крім того, існують

численні варіації методу зворотного поширення помилки, розроблені з метою збільшення швидкості протікання процесу навчання.

Також варто відзначити, що одношарова нейронна мережа істотно обмежена в тому, навчання яким шаблонами вхідних даних вона підлягає, в той час, як багатошарова мережа (з одним або більше прихованим шаром) не має такого недоліку. На рис.2.21 представлена типова структура мережі, що складається з одного вхідного шару, одного прихованого шару і вихідного шару з 5-и нейронів, яка може обслуговувати 5 об'єктів.

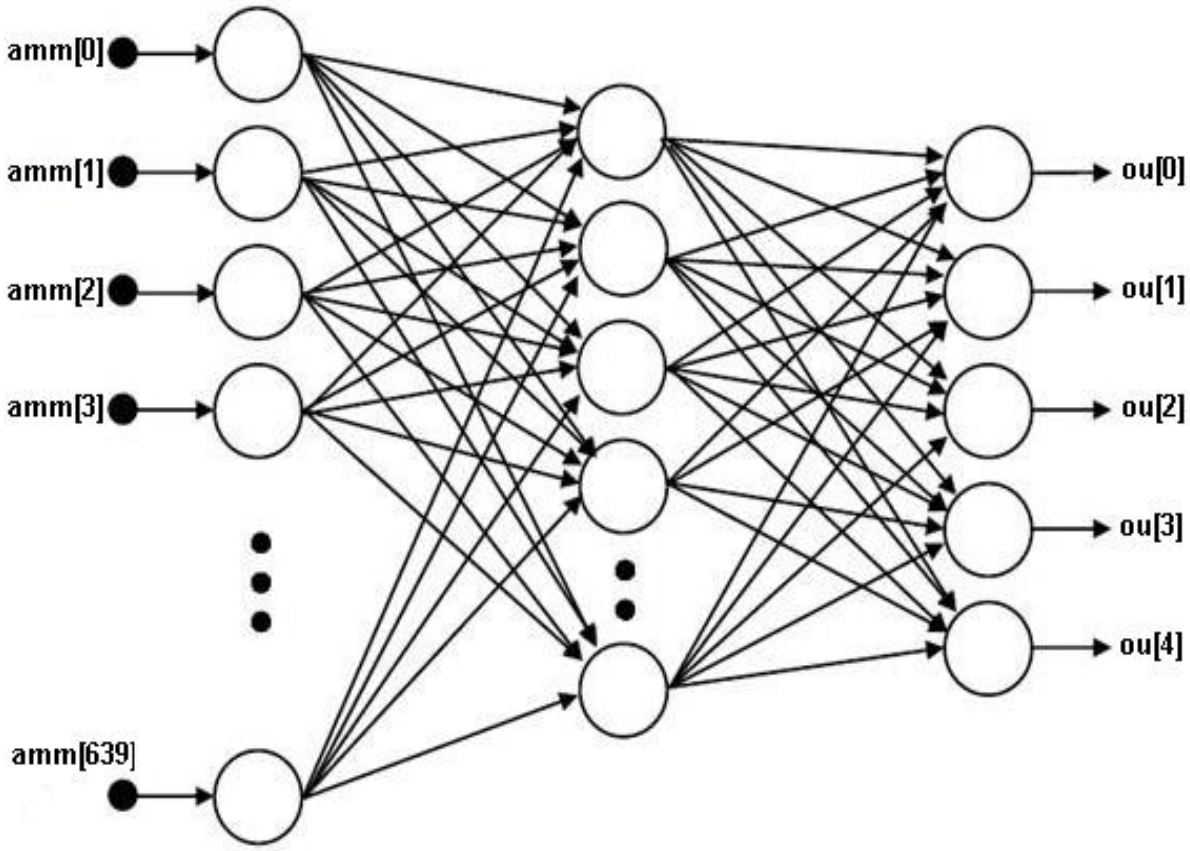


Рис.2.21. Структура штучної нейронної мережі

Кількість нейронів вхідного шару має дорівнювати числу елементів масиву, що характеризує слово-команду. Представлена на рис.2.21 структура є типовою для створення моделі мережі відповідно до бібліотеки FANN.

За умовою завдання представлена на рис.2.7 структура повинна мати 640 нейронів у вхідному шарі, 7 нейронів у вихідному. Експерементально було

виявлено, що мережа повинна мати 2 прихованих шари по 100 нейронів. Навчання нейронної мережі виконується програмою n.php [4].

Опис алгоритму програми:

1. Зчитує значення 42-масивів з 8-ми каталогів. У каталогах з 0 по 6 знаходяться 42 масива, які сформовані з слів-команд синій, червоний, зелений, класний, сильний, земля, температура. Кожне слово вимовлялося 42 рази різною інтонацією і з різним становищем мікрофона. У 7-му каталозі знаходиться 42 масива, які описують інші слова.

2. Формує масив даних для навчання нейронної мережі. Він являє собою перерахування наборів даних і значень, які повинні бути на виході мережі. Більш детально це представлено коментарями в програмі n.php.

3. $ Ann = fann_create (array (640, 100, 100, 7), 1.0, 0.7). тут

640 - кількість вхідних даних (640 нейронів вхідного шару);

100 - кількість нейронів 2-го прихованого шару;

100 - кількість нейронів 3-го прихованого шару;

7 - кількість вихідних нейронів;

1.0 - Можливості підключення нейромережі (1 – повнозвязана для персептрона);

0.7 - Параметр, що дозволяє управляти величиною корекції ваг на кожній ітерації (в алгоритмі зворотного поширення помилки він вводить, як коефіцієнт при градієнті).

4. Навчає мережу за допомогою функції fann_train ():

fann_train ($ ann, $ my, 1000, 0.0001, 10).

тут

1000 - кількість ітерацій при навчанні мережі;

0.01 - допустима похибка;

10 - проміжки, через які виводиться звіт про навчання.

5. Зберігає навчену мережу в файлі "my.ann" для подальшого використання:

fann_save ($ ann, "my.ann").

Розпізнавання слова виконується програмою ru.php [4]. Алгоритм програми:

1. Завантажує навчену модель мережі з файлу "my.ann":

$ Ann = fann_create ("my.ann");

2. Зчитує з файлу "r1.txt" масив $ amm [], який визначає слово.

3. Запускає функцію fann_run ($ ann, $ amm) для визначення, яким словом відповідає лічений масив (тобто. Виконує розпізнавання). Виводить розпізнане слово - команду.

4. Виконує передачу даних коду розпізнаної команди на контролер Ардуіно через Bluetooth. Наприклад, команда "червоний" відповідає символу "1" (див. Програму ru.php). Після отримання цього символу Ардуіно виконує включення відповідного пристрою. Команда "температура" посилає символ "6". Після отримання його Ардуіно виконує опитування температурного датчика, і значення температури посилає назад через Bluetooth комп'ютера. Програма temp.php [4] постійно прослуховує Bluetooth Ардуіно і після отримання температури записує її значення в файл "aa.a". Програма ru.php зчитує файл "aa.a" і роздруковує його на моніторі. Слід враховувати, що програма temp.php повинна бути запущена на інший консолі.

Схема підключення Arduino до Bluetooth, виконавчих пристроїв і датчику температури, представлені в наступному розділі.

Для запуску системи розпізнавання і виконання команд необхідно запустити командний файл. Він працює в циклі кожні 3 секунд після чергового розпізнавання і виконання команди:

```
#!/bin/sh
while (true)
do
rm aa.a
echo
echo "ГОВОРІТЬ"
arecord -q -d 2 -f cd -r 16000 -c 1 a.wav
```

```
./slice
./new
./ru.php
sleep 3
done
```

Перед його запуском необхідно підключити Bluetooth командою sudo rfcomm bind/dev/rfcomm1 98: D3: 31: B0: 86: 16, 1 (тут використовується адреса експериментального Bluetooth пристрою, підключеного до Ардуїно) а в іншій консолі запустити програму temp.php. Якщо будуть виникати помилки щодо зайнятості пристрою, то необхідно перезапустити програму temp.php.

Висновки за розділом 2

1. Показана можливість надійного розпізнавання слів за допомогою ДПФ інтервалів слів довжиною 15...23МС з виділенням 8-ми локальних максимумів амплітуд. Відновлення слова за допомогою функцій синуса показало подібність звуку вихідного слова до 90%.

2. За допомогою бібліотеки FANN математичної моделі нейронної мережі показана практична можливість побудови нейронної мережі розпізнавання 7-ми команд з можливим масштабуванням.

3. Представлений підхід практичного вирішення задачі розпізнавання команд можна використовувати для розпізнавання декількох десятків команд і відповідно управління декількома десятками пристроями через контролер Arduino.

4. Представлене тут рішення можна використовувати для створення систем управління голосом на будь-якій мові, будь-якими звуками.

5. Помічені недоліки:

5.1. Залежність якості розпізнавання від диктора, мікрофона і навіть розташування мікрофона щодо диктора. Для вирішення цих проблем необхідно створення великої бази даних слів, вимовлених різними дикторами з різними мікрофонами і при різних розташуваннях мікрофонів.

5.2. Для надійного (практично 100% правильну вимову слова) розпізнавання слів необхідно відсутність сторонніх звуків під час вимови слова. В основному сторонні шуми впливають на виділення слова з 2-х секундного аудіо файлу a.wav програмою slice.c. Можна припустити, що якщо виділення слова при наявності шумів буде правильним, то нейронна мережа для різних варіантів слів з шумами зможе правильно виконати розпізнавання (аналогія розпізнавання людиною слів з урахуванням перешкод).

5.3. При великій кількості слів необхідно формувати досить громіздкий масив даних для навчання нейронної мережі при використанні бібліотеки FANN (див. Програму n.php). Можна припустити що це є основним обмеженням використання бібліотеки FANN методом розпізнавання за словами.

6. Порівняння похибок розпізнавання слів з використанням нейронної мережі у залежності від алгоритму існуючих методів розпізнавання табл. 2.1.

Таблиця 2.1

Результати досліджень

Система	Міра помилки, %
Контекстно незалежна ПММ	38.85
Контестно залежна ПММ	35.21
BLSTM/HMM	33.84
BLSTM/HMM із зваженими помилками	31.57
HTK (метод пошуку кращого шляху)	31.47
HTK (метод пошуку кращого префіксу)	30.51

РОЗДІЛ ІІІ. КОНСТРУКТОРСЬКА ЧАСТИНА

3.1. Огляд проектування

В ході даної магістерської дисертації, поставлена задача - спроектувати пристрій розпізнавання мови для голосового управління кліматом в приміщенні з високоточними показниками розпізнавання голосу. Важливо розуміти, що завдання створення приладу відноситься більше до розпізнавання мови, оскільки завдання дисертації досягти найвищих показників розпізнавання голосу. Мета створення приладу голосового управління складається з декількох пунктів. Її постановка включає в себе:

а) Ознайомлення з матеріалом і прикладами скетчів в середовищі Arduino за темами голосового управління.

б) Вивчення принципових схем існуючих схожих проектів.

в) Складання власної принципової схеми.

г) Вибір відповідних комплектуючих.

д) Пошук і доступ до потрібних комплектуючих для пристрою.

Для досягнення поставленої мети обрані плати на основі контролера ATmega328. На рис.3.1., представлений сам мікроконтролер ATmega328.

Рис.3.1. Мікроконтролер ATmega328

Дані мікроконтролери широко застосовуються в сфері програмування. Їх основними перевагами є хороші технічні характеристики і габарити, які повністю відповідають світовим стандартам сучасної техніки.

В табл 3.1 представлені технічні характеристики даного мікроконтролера.

Таблиця 3.1

Технічні характеристики мікроконтролера ATmega328

Найменування	Значення
Тактова частота, МГц	20
Обсяг Flash-пам'яті, КБ	32
Обсяг SRAM-пам'яті, КБ	2
Обсяг EEPROM-пам'яті, КБ	1
Напруга живлення, В	1.8-5.5
Загальна кількість портів	23
кількість ШІМ	6
кількість АЦП	6
дозвіл АЦП	10 біт

Існує широкий спектр плат arduino для роботи з даними контролером.

Arduino - це інструмент для проектування електронних пристроїв (електронний конструктор) який більш щільно взаємодіє з навколишнім фізичним середовищем, ніж стандартні персональні комп'ютери, які фактично не виходять за рамки віртуальності. Це платформа, призначена для «physical computing» з відкритим програмним кодом, побудована на простій друкованої платі з сучасним середовищем для написання програмного забезпечення.

Arduino застосовується для створення електронних пристроїв з можливістю прийому сигналів від різних цифрових і аналогових датчиків, які можуть бути підключені до нього, і управління різними виконавчими

пристроями. Проекти пристроїв, засновані на Arduino, можуть працювати самостійно або взаємодіяти з програмним забезпеченням на комп'ютері (напр.: Flash, Processing, MaxMSP). Плати можуть бути зібрані користувачем самостійно або куплені в зборі. Середовище розробки програм з відкритим вихідним текстом доступна для безкоштовного скачування.

Кожна плата arduino має певну кількість цифрових і аналогових виходів. Також багато плат arduino мають свою варіацію мікроконтролера ATmega328. Виходи платформи arduino можуть працювати як входи або як виходи.

Даний документ пояснює функціонування виходів в цих режимах. Також необхідно звернути увагу на те, що більшість аналогових входів arduino можуть конфігуруватись і працювати так само як і цифрові порти введення/виводу.

Виходи arduino стандартно налаштовані як порти введення, таким чином, не потрібно явної конфігурації в функції pinMode().

Сконфігуровані порти введення знаходяться в високо-імпедансному стані. Це означає те, що порт введення має мале навантаження на схему, в яку він включений. Еквівалентом внутрішньому опору буде резистор 100 МОм підключений до виходу мікросхеми. Таким чином, для переходу порту введення з одного стану в інший потрібно маленьке значення струму. Це дозволяє застосовувати виходи мікросхеми для підключення ємнісного датчика, фотодіода, аналогового датчика зі схемою, схожою на RC-ланцюг.

Якщо до даного виходу нічого не підключено, то значення на ньому буду прийматись як випадкові величини, що наводяться електричними перешкодами або ємнісним взаємозв'язком з сусіднім виходами.

Виходи, сконфігуровані як порти виводу, знаходяться в низько-імпедансном стані. Дані виводи можуть пропускати через себе досить великий струм. Виводи мікросхеми Atmega можуть бути джерелом (позитивний) або приймачем (негативний) струму до 40 мА для інших пристроїв. Такого значення струму досить, щоб підключити світлодіод (обов'язковий послідовно

включений резистор), датчики, але недостатньо для більшості реле, соленоїдів і двигунів.

Коротке замикання виводів Arduino або спроби підключити енергоємні пристрої можуть пошкодити вихідні транзистори виведення або весь мікроконтролер Atmega. У більшості випадків дані дії приведуть до відключення виводів на мікроконтролері, але інша частина схеми буде працювати згідно з програмою. Рекомендується до виходів платформи підключати пристрої через резистори 470 Ом або 1 кОм, якщо пристрою не потрібно більший струм для роботи.

Мікроконтролери Atmega, використовувані в Arduino, містять шестиканальний аналого-цифровий перетворювач (АЦП). Дозвіл перетворювача складає 10 біт, що дозволяє на виході отримувати значення від 0 до 1023. Основним застосуванням аналогових входів більшості платформ Arduino є зчитування аналогових сигналів датчиком, але в той же час вони мають функціональність вводів/виводів широкого застосування (GPIO) (те ж, що і цифрові порти введення/виводу 0 - 13).

Таким чином, при необхідності застосування додаткових портів введення/виводу є можливість конфігурувати невикористовувані аналогові входи.

Для виводу, який працював раніше як цифровий порт виводу, команда analogRead буде працювати некоректно. В цьому випадку рекомендується налаштувати його як аналоговий вхід. Аналогічно, якщо висновок працював як цифровий порт виведення зі значенням HIGH, то зворотна установка на введення підключить підтягуючий резистор.

Керування на мікроконтролер Atmega не рекомендує проводити швидке переключення між аналоговими входами для їх читання. Це може викликати накладення сигналів і створити спотворення в аналогову систему. Однак після роботи аналогового входу в цифровому режимі потрібно налаштувати паузу між читанням функцією analogRead() та інших входів.

3.2 Вибір елементів

3.2.1 Вибір мікроконтролера.

Для виконання поставленого завдання технічні характеристики задовольняють лише деякі плати arduino на платформі ATmega 328.

ArduinoNano має найбільш компактні розміри, що робить роботу з платою дуже зручною (рис. 3.2)

Рис. 3.2 Плата ArduinoNano

Дану плату можна вибрати виходячи з зручних розмірів. Однак плата має порівняно невелику кількість цифрових і аналогових входів і виходів, меншу пам'ять, але найбільшу зручність при роботі. В табл. 3.2 вказані технічні характеристики.

Таблиця 3.2

Технічні характеристики плати ArduinoNano

Найменування	Значення
Робоча напруга	5 В
Вхідна напруга	7-12 В
Цифрові входи/виходи	14 (6 на вихід ШІМ)
Аналогові входи	6
Флеш пам`ять	16 КБ

ОЗУ	1 КБ
EEPROM	512б
Тактова частота	16 Мгц
Споживчий струм	40мА

ArduinoNano має 14 цифрових входів і виходів 6 з яких йдуть на вихід ШІМ. Також плата має 6 аналогових входів. Вибір ArduinoNano дозволить створити досить продуктивний пристрій з мінімальними габаритними розмірами, що є досить актуальним питанням у розвитку сучасних технологій. На рис.3.3. представлена схема всіх входів і виходів плати ArduinoNano.

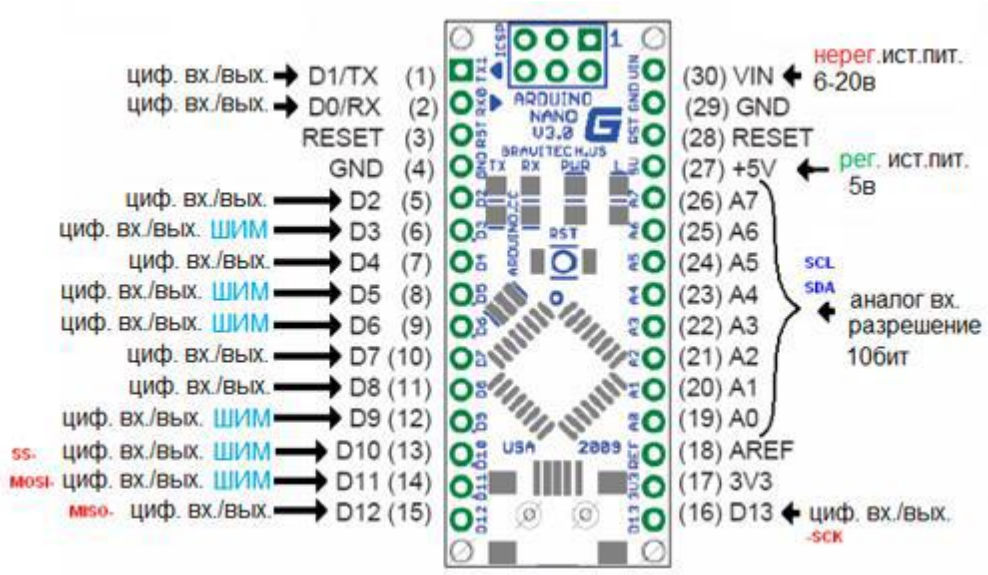


Рис.3.3. Входи і виходи ArduinoNano

ArduinoMega-плата, яку так само можна використовувати при виконанні поставленого завдання (рис.3.4.).

Дану плату можна вибрати в зв'язку з високими технічними характеристиками, кількістю входів і виходів. ArduinoMega має найбільшу кількість цифрових і аналогових входів і виходів, велику пам'ять, проте і габарити плати дещо більше аналогів.

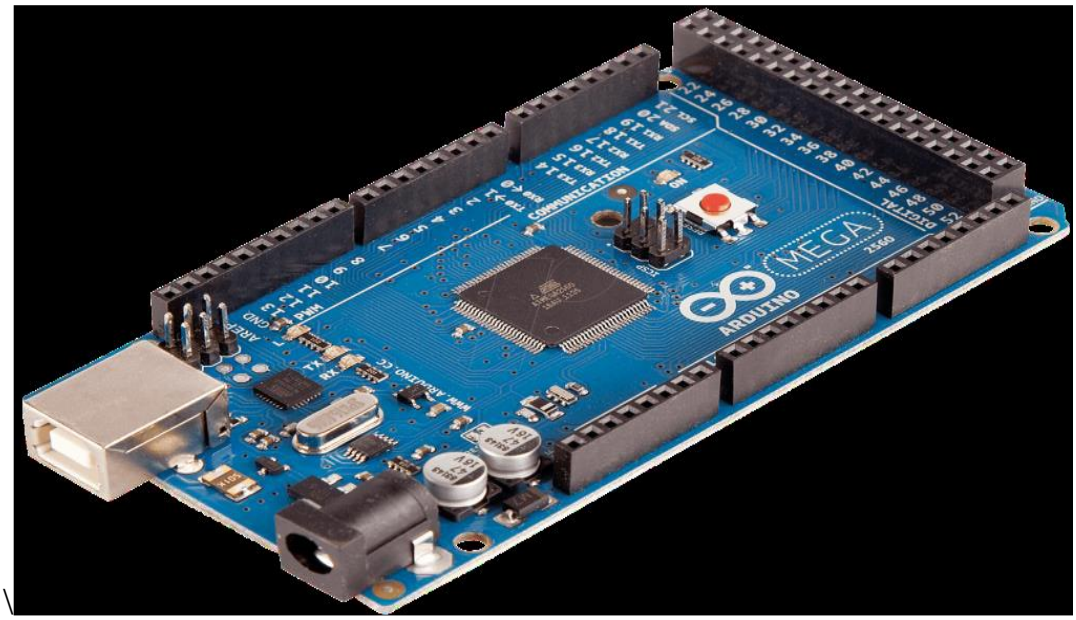


Рис.3.4.– плата ArduinoMega

Дана плата дуже зручна в роботі і дозволяє проектувати на її основі найскладніші пристрої які виконують цілий ряд функцій. Харчування і зв'язок з комп'ютером здійснюються через UART інтерфейс. В табл. 3.3 представлені технічні характеристики ArduinoMega.

Таблиця 3.3.

Найменування	Значення
Робоча напруга	5 В
Вхідна напруга	7-12 В
Цифрові входи/виходи	54 (14 на вихід ШІМ)
Аналогові входи	16
Флеш-пам'ять	256 КБ
ОЗУ	8 КБ
EEPROM	4 КБ
Тактова частота	16 МГц
Споживчий струм	40мА

ArduinoMega має 54 цифрових входів і виходів, 14 з яких йдуть на вихід ШІМ. Так само плата має 16 аналогових входів. Вибір ArduinoMega дозволить створювати складні пристрої, відповідальні за цілий ряд функцій, так само пам'ять плати дозволяє запам'ятати складні скетчі з великим обсягом. На рис.3.5 представлена схема з входами та виходами ArduinoMega.

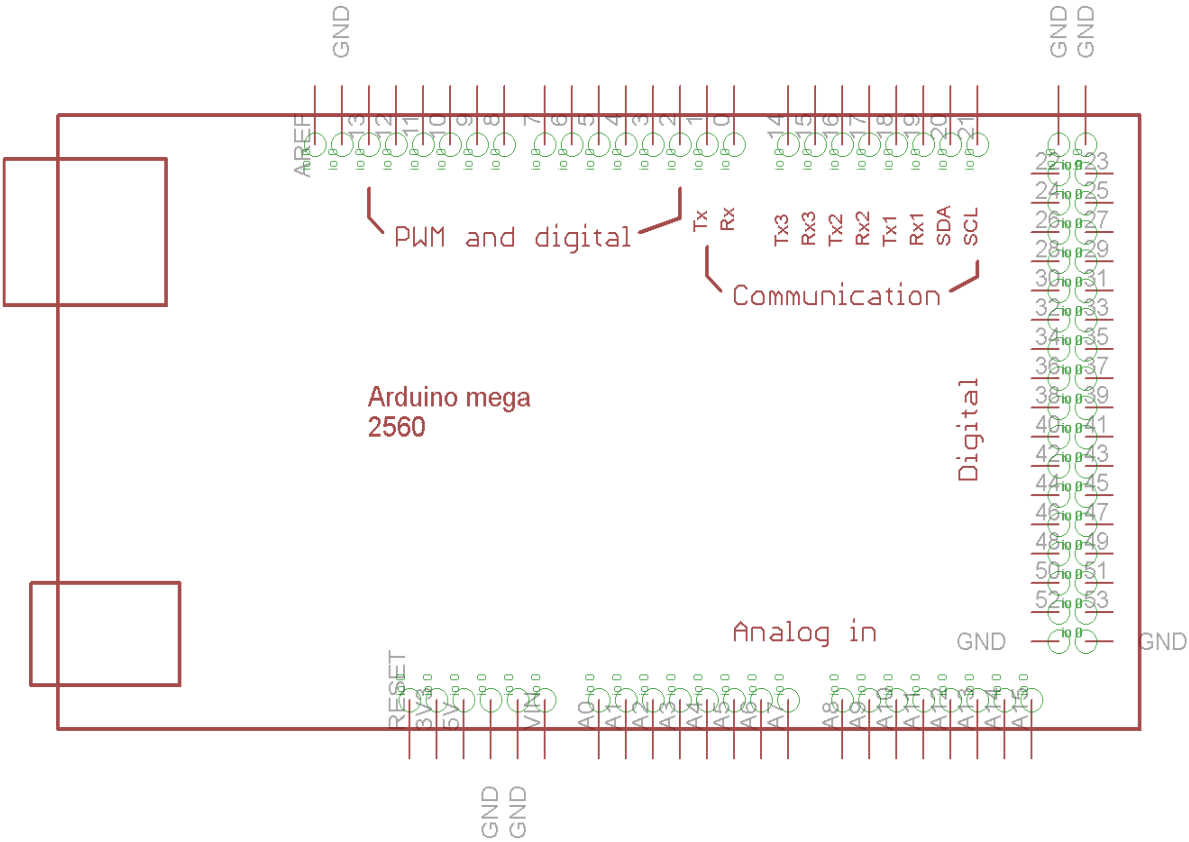


Рис.3.5. - Входи і виходи ArduinoMega

В результаті, для поставленої задачі був вибраний мікроконтролер ArduinoUno на платформі AtmelATmega328 (рис.3.6.)

На вибір саме цього контролера вплинули його технічні дані, більша кількість пам'яті, входів та виходів в порівнянні з ArduinoNano і більш компактні розміри в порівнянні з ArduinoMega

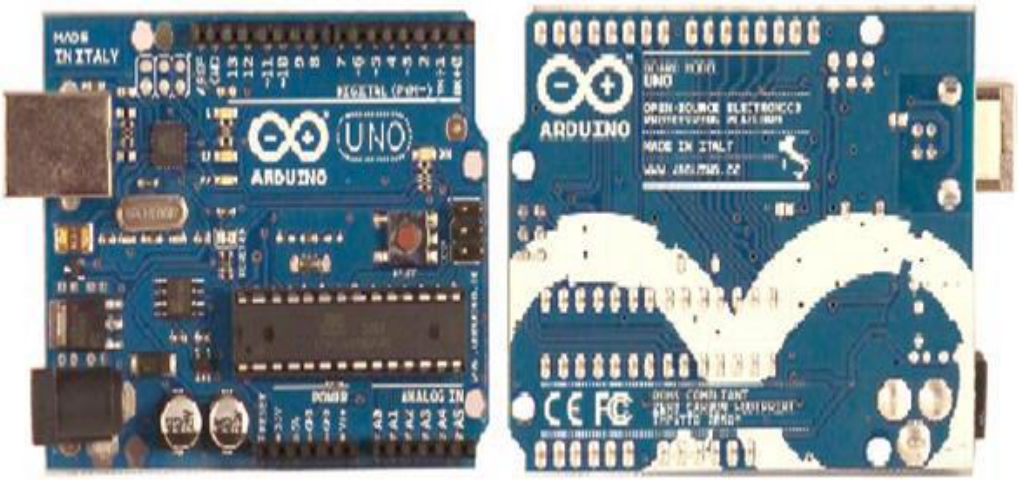


Рис.3.6 – Плата ArduinoUno

В табл. 3.4 вказані технічні характеристики плати мікроконтролера.

Таблиця 3.4

Найменування	Значення
Робоча напруга	5 В
Вхідна напруга	7-12 В
Цифрові входи/виходи	14 (6 на вихід ШІМ)
Аналоговий входи	6
Флеш пам`ять	32 КБ
ОЗУ	2 КБ
EEPROM	1 КБ
Тактова частота	16 МГц
Споживчий струм	40мА

Плата має невеликі розміри, тому зручна у використанні і монтажі. Вибір даної плати користується у програмістів найбільшою популярністю, оскільки відповідає всім вимогам користувачів. На платі ArduinoUNO є мікроконтролер. На рис.3.7 представлено креслення плати із зазначенням розмірів в міліметрах.

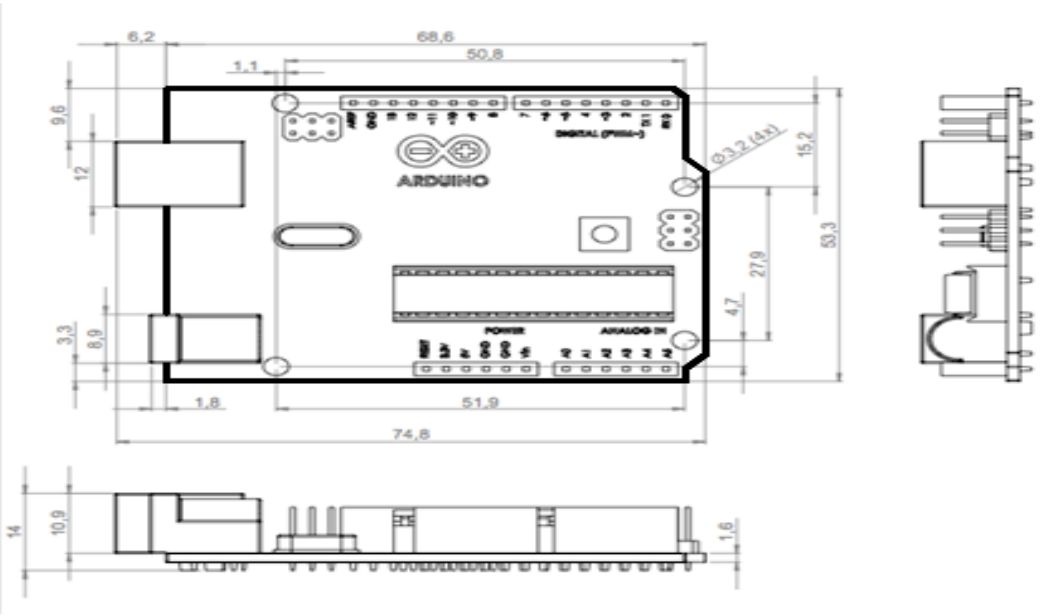


Рис.3.7. - Креслення плати Arduino Uno

Кожен з 14 цифрових виходів плати може використовуватися в якості введення або виведення. Крім того деякі виводи мають особливі функції. Контакти 0 (RX) і 1 (TX) використовуються для передачі і отримання даних. Виводи 2 і 3 можуть бути налаштовані на виклик переривання. Контакти 3,5,6,9,10 і 11 забезпечують 8-бітний ШІМ вихід. Так само є послідовний периферійний інтерфейс SPI, контакти 10 (SS), 11 (MOSI), 12 (MISO), 13 (SCK).

На платформі Nano встановлено 6 аналогових входів, кожен дозволом 10 біт (тобто може приймати 1024 різних значення). Стандартно виводи мають діапазон вимірювання до 5 В відносно землі. На рис.3.8 представлені входи і виходи плати із зазначенням всіх виводів і розмірів в міліметрах.

ArduinoUno може отримувати живлення через підключення USB, або від нерегульованого 6-20 В, або регульованого 5 В зовнішнього джерела живлення. Автоматично вибирається джерело з найвищим напругою.

На платформі ArduinoUno встановлені деякі пристрої для здійснення зв'язку з комп'ютером, іншими Arduino або мікроконтролерами. ATmega328 підтримує послідовний інтерфейс UART TTL (5 В), здійснюваний виводами 0 (RX) і 1 (TX). Встановлена на платі мікросхема ATmega8U2 направляє даний

інтерфейс через USB, а програми з боку комп'ютера «спілкуються» з Arduino через віртуальний COM порт. Прошивка ATmega8U2 використовує стандартні драйвера USB COM, не вимагаючи сторонніх драйверів, однак для ОС Windows для правильного підключення потрібно файл ArduinoUNO.inf.

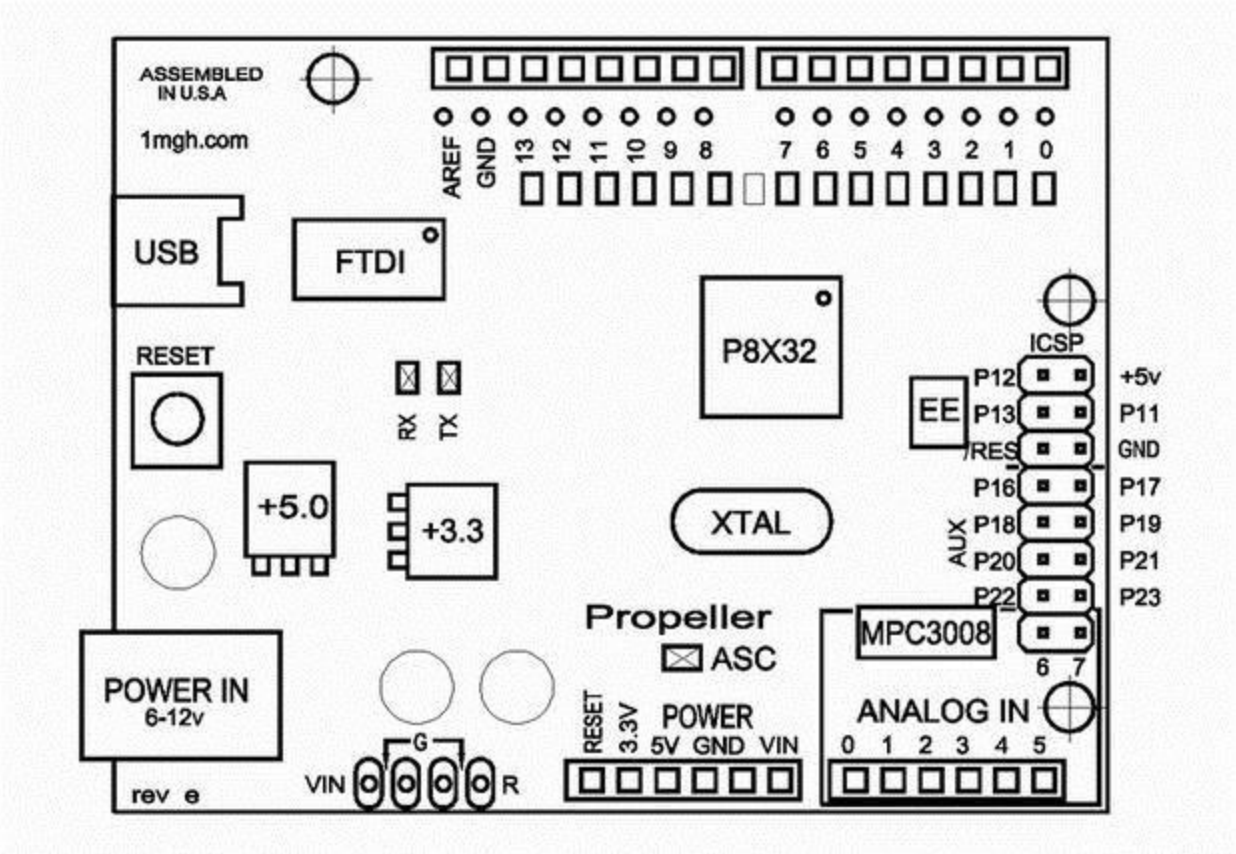


Рис.3.8. - Контакти ArduinoUno

Моніторинг послідовної шини (SerialMonitor) програми Arduino дозволяє посилати і отримувати текстові дані при підключенні до платформи. Світлодіоди RX і TX на платформі будуть мигати при передачі даних через мікросхему FTDI або USB підключення (але не при використанні послідовної передачі через висновки 0 і 1). Крім установки зв'язку між комп'ютером і платою, буде потрібно додаток бібліотеки SoftwareSerial [12].

3.3 Середовище для проведення дослідів

3.3.1 Принципова схема взаємодії з приладами

Для проведення перевірки точності голосового керування використовувались схеми наведені на рис.3.9 та рис.3.10.

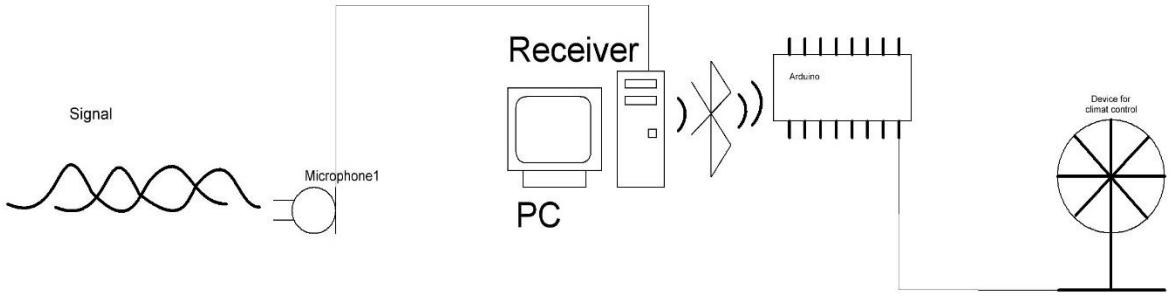


Рис.3.9 Схема проведення досліджень в точності розпізнавання голосу

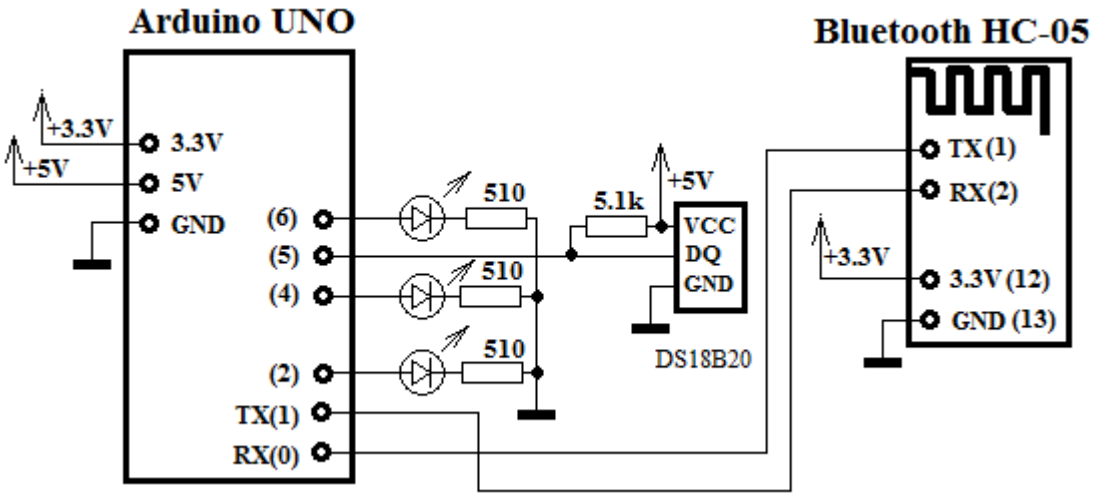


Рис.3.10. Схема підключення пристроїв до Arduino

Введення команд виконувалось через стандартний мікрофон, підключений через популярний аудіо адаптер CMI 8738/PCI (рис.3.11) до комп'ютера, який працює під управлінням операційної системи Linux Ubuntu 10.04.

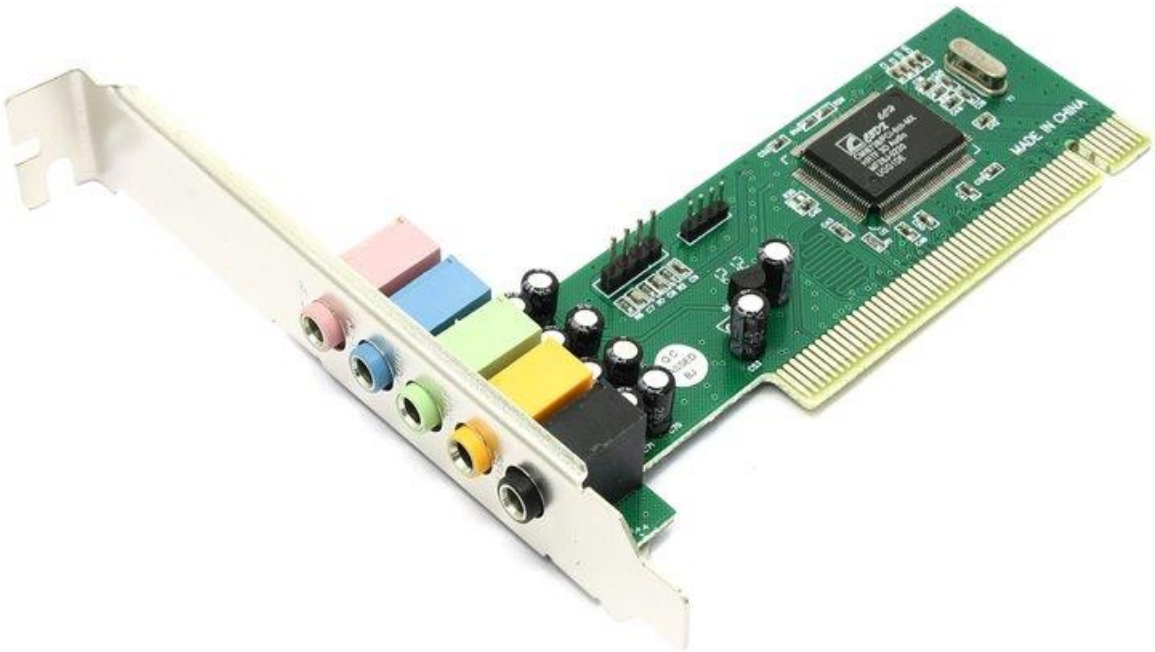


Рис.3.11. Стандартна звукова карта CMI 8738/PCI

В якості приймача звуку було обрано стандартний мікрофон фірми Sven(рис.3.12).В табл. 3.1 наведено його характеристики.

Рис.3.12. Мікрофон моделі Sven MK-200

Таблиця 3.1

Роз'єм	Mini-jack (3.5 мм)
Чутливість	-60 ± 3 дБ
Діапазон	50 - 16000 Гц
Довжина кабеля	1.8 м
Вага	63 г

На рис.3.10 показана схема підключення до Arduino UNO Bluetooth датчика HC-05, датчика температури DS18B20 і трьох світлодіодів червоного, зеленого і синього кольору (альтернатива трьом виконавчим механізмам, наприклад світло, телевізор і кондиціонер).

Програма для Arduino написана в стандартному середовищі розробки Arduino:

#include <OneWire.h> // Підключаємо бібліотеки шини OneWire

#include <DallasTemperature.h> // Підключаємо опис бібліотеки для визначення температури (DS18B20)

#define ONE_WIRE_BUS 5 // Датчик температури підключений до 5-го виводу Ардуіно

char inByte; // вхідні дані

int RED = 2; // RED підключений до 2 виводу

int GR = 4; // GREEN підключений до 4 виводу

int BL = 6; // BLUE підключений до 6 виводу

OneWire oneWire (ONE_WIRE_BUS); // Налаштування шини 1wire для роботи з 5-м висновком Ардуіно

DallasTemperature sensors (& oneWire); // Підключаємо датчик температури

```
void setup () {
  Serial.begin (9600); // ініціалізація порту
  sensors.begin (); // Ініціалізація датчика температури DS18B20
  pinMode (RED, OUTPUT); // Установка 2-го виведення на вихід
  pinMode (GR, OUTPUT); // Установка 4-го виведення на вихід
  pinMode (BL, OUTPUT); // Установка 6-го виведення на вихід
  sensors.requestTemperatures (); // Запит температури
  int temp = sensors.getTempCByIndex (0); // Отримання температури з нульового датчика
}
void loop () {
  if (Serial.available ()> 0) {// якщо прийшли дані
    inByte = Serial.read (); // зчитуємо байт
    if (inByte == '0') {
      digitalWrite (RED, LOW); // якщо 0, то вимикаємо RED
    }
    if (inByte == '1') {
      digitalWrite (RED, HIGH); // якщо 1, то включаємо RED
    }
    if (inByte == '2') {
      digitalWrite (GR, LOW); // якщо 2, то вимикаємо GREEN
    }
    if (inByte == '3') {
      digitalWrite (GR, HIGH); // якщо 3, то включаємо GREEN
    }
    if (inByte == '4') {
```

```
    digitalWrite (BL, LOW); // якщо 4, то вимикаємо BLUE
   }
   if (inByte == '5') {
    digitalWrite (BL, HIGH); // якщо 5, то включаємо BLUE
   }
   if (inByte == '6') {// якщо 6, то зчитуємо температуру і посилаємо її на
bluetooth
      sensors.requestTemperatures ();
      int temp = sensors.getTempCByIndex (0);
      Serial.print ( "Температура,");
     Serial.print (temp); Serial.print ( "");
   }
  }
 }
```

Висновки за розділом 3

В даному розділі проведено аналіз досупних мікроконтроллерів і плат для реалізації голосового управління, та обрано отсаточний варіант.

Були представлені схеми під'єднання пристроїв які взаємодіють за допомогою комп'ютера та Bluetooth датчика HC-05.

Показана принципова схема підключення виконавчих механізмів.

Програма для Arduino написана в середовищі в середовищі розробки Ардуіно на C ++ (проект Wiring).

РОЗДІЛ IV. СТВОРЕННЯ СТАРТАП ПРОЕКТУ

Ціль проекту створення електронного пристрою розпізнання голосу для керування кліматом в приміщенні який володіє високими показниками точності в розпізнаванні голосових команд .

Поставлені задачі

- Використання прогресивного методу враховуючі недоліки існуючих рішень.
- Створення максимально простого і доступного у використанні приладу.
- Спростити доступ до техніки і роботи з нею.
- Урахування габаритності і мобільності пристрою.
- Розробка ідентифікації користувача по голосу.
- Доступна ціна для кожного.
- Можливість включення пристрою голосового управління для інших цілей.

Опис ринку

Загальною проблемою на ринку подібних систем є низькі показники точності, шумоподавлення, висока вартість та складна мобільність пристрою.

Актуальність

а) Збільшення попиту на програми розпізнавання мови на мобільних пристроях.

б) Зростання попиту на послуги голосової аутентифікації для мобільного банкінгу.

в) Інтеграція голосової верифікації і розпізнавання мови.

Фінансовий аналіз

Загальна вартість комплектуючих пристрою 2 138 грн, в порівнянні з іншими пристроями ця сума не значна.

Споживач

Продукт орієнтований на:

1. Широке коло людей котрі користуються сучасними пристроями.
2. Людей з обмеженими можливостями.
3. Посередники на ринку.

Ринок збуту

У перший рік існування стартап орієнтується на регіональних (Київська область) та українських споживачів, після чого плануються поставки за кордон України (здебільшого країни ближнього зарубіжжя).

Конкурентні переваги

Показники, за якими продукт виграє на ринку:

1. Ціна
2. Портативність
3. Простота обслуговування

Показники, що не витримують конкуренції:

1. Точність вимірювання
2. Універсальність застосування.

Техніко-економічні показники

Капіталовкладення: 200000 грн

Ринкова вартість одиниці продукції: ~3500 грн

Собівартість одиниці продукції: 2500 грн

Прибуток з одиниці продукції: 1000 грн

Запланований випуск продукції на перший рік існування стартапу: 1000 одиниць в табл. 4.1.

Таблиця 4.1.

Запланований випуск продукції

Місяць	Червень 2018	Липень 2018	Серпень 2018	Вересень 2018	Жовтень 2018	Листопад 2018	Грудень 2018
Запланований обсяг випуску	50	50	50	50	50	50	100

Час повернення капіталовкладень: 4 місяці

Рентабельність: 28.5%

Висновки за розділом 4

В результаті проведеного маркетингового аналізу перспектив реалізації запропонованих науково-технічних рішень та пропозицій, оцінювання можливостей їх ринкового впровадження можна стверджувати, що розроблюваний проект має можливість ринкової комерціалізації та може бути рентабельним проектом на ринку. Зростання попиту на аналогічні товари додає масовості придбання подібних пристроїв, але створює жорсткі конкурентні умови виходу на ринок.

Проект має високі перспективи впровадження з огляду на сучасний стан промисловості, яка потребує нових потужних та економних рішень. Бар'єрами входження на ринок може бути відсутність масового виробника, сильний конкурентний тиск з боку великих фірм аналогічних продуктів, потреба у великій кількості кваліфікованих кадрів та дорогої точної апаратури. Але якщо

правильно розставити пріоритети, зарекомендувати себе на ринку і грамотно вести бізнес, то проект має великі шанси на ріст та гідний прибуток.

Подальша імплементація проекту є доцільною та рентабельною.

ВИСНОВКИ

У дисертаційній роботі вирішено актуальне наукове завдання – розробка пристрою розпізнавання голосу з високими показниками точності в розпізнаванні голосу за допомогою використання рекурентних нейронних мереж.

Основні наукові теоретичні та практичні результати полягають в наступному:

1. Проведений огляд процесу розпізнавання голосу на основі існуючих методів, що дозволило виявити переваги та недоліки існуючих методів та виділити особливості процесу розпізнавання мови, на які необхідно звертати увагу під час проектування пристрою.
2. Найбільш перспективним вбачається варіант побудови пристрою ГК на основі розпізнання голосу за допомогою РНМ. Це дозволить досягти високої точності та завадостійкості в розпізнаванні голосових команд.
3. Проведений огляд відомих принципів побудови нейронних мереж та способів їх реалізації що дозволило виявити переваги та недоліки існуючих рішень та виділити рішення, які можуть розглядатись як прототипи.
4. Побудова пристрою розпізнавання мови на основі РНМ забезпечує можливість збільшення голосових команд шляхом простого додавання даних в базу НМ, що є дуже істотним, оскільки підключення багатьох пристроїв є суттєвою перевагою.
5. Спільне використання рекурентних НМ та перетворення Фур’є дозволяє поєднати переваги обох способів: забезпечити високу точність відтворення заданого сигналу в разі необхідності.
6. Запропоновано методику отримання масивів даних з отриманого сигналу.

7. У результаті дослідження було отримано результати розпізнання команд при визначеній базі даних, які дозволяють стверджувати, що при записі одного слова 42 рази в різних умовах, в подальшому сприяють на точність розпізнавання.

8. Було спроектовано схема при якій вимовляння голосових команд приводило в дій прилади.

9. Експериментальні дослідження показали, що математичні розрахунки та імітаційне моделювання, проведені у попередніх розділах були виконані коректно.

10. В результаті проведеного маркетингового аналізу перспектив реалізації запропонованих науково-технічних рішень та пропозицій, оцінювання можливостей їх ринкового впровадження можна стверджувати, що розроблюваний проект має можливість ринкової комерціалізації та може бути рентабельним проектом на ринку. Зростання попиту на аналогічні товари додає масовості придбання подібних пристроїв, але створює жорсткі конкурентні умови виходу на ринок.

11. Проект має високі перспективи впровадження з огляду на сучасний стан ринку, який потребує нових потужних та економних рішень. Бар’єрами входження на ринок може бути відсутність масового виробника, сильний конкурентний тиск з боку великих фірм аналогічних продуктів, потреба у великій кількості кваліфікованих кадрів та дорогої точної апаратури. Але якщо правильно розставити пріоритети, зарекомендувати себе на ринку і грамотно вести бізнес, то проект має великі шанси на ріст та гідний прибуток. Подальша імплементація проекту є доцільною та рентабельною.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Ардуїно [Електронний ресурс] // Arduino.ru. – 2015. – Режим доступу до ресурсу: http://arduino.ru/.
2. Перспективи на ринку систем голосового управління [Електронний ресурс] // Хабрахарбр. – 2016. – Режим доступу до ресурсу: https: //habrahabr.ru/post/232613/.
3. Розпізнавання мови. [Електронний ресурс] // Вікіпедія. – 2007. – Режим доступу до ресурсу: https://ru.wikipedia.org/wiki/Распознавание_речи.
4. Голосове управління [Електронний ресурс] // Вікіпедія. – 2007. – Режим доступу до ресурсу: https://ru.wikipedia.org/wiki/Голосовое_управление.
5. Уллі С. Программирование микроконтроллерных плат Arduino/Freeduino / Соммер Уллі. – Петербург, 2012.
6. Ревич Ю. Цікава електроніка / Юрій Ревич. – Петербург, 2015.
7. Карвинен Т. Робимо сенсори. Проекти сенсорних пристроїв на базі Arduino і Raspberry Pi / Т. Карвинен, К. Карвинен, В. Валтокарі., 2015.
8. Петрін В. О. Проекти з використанням контролера Arduino. 2 изд. / Віктор Олександрович Петрін..
9. Голосове управління Arduino засобами Processing і Google Speech API [Електронний ресурс]. – 13. – Режим доступу до ресурсу: https://habrahabr.ru/post/236673/.
10. Голосове управління вимикачами на Arduino [Електронний ресурс]. – 30. – Режим доступу до ресурсу: http://compcar.ru/forum/showthread.php?t=8016.
11. Перетворення Лапласа [Електронний ресурс] // Вікіпедія – Режим доступу до ресурсу: https://ru.wikipedia.org/wiki/%D0%9F%D1%80%D0%B5%D0%BE%D0%B1%D1%80%D0%B0%D0%B7%D0%BE%D0%B2%D0%B0%D0%BD%D0%B

8%D0%B5_%D0%9B%D0%B0%D0%BF%D0%BB%D0%B0%D1%81%D0%B0

12. Частота дискретизації [Електронний ресурс] // ВІкіпедія – Режим доступу до ресурсу

:https://uk.wikipedia.org/wiki/%D0%A7%D0%B0%D1%81%D1%82%D0%BE%D1%82%D0%B0_%D0%B4%D0%B8%D1%81%D0%BA%D1%80%D0%B5%D1%82%D0%B8%D0%B7%D0%B0%D1%86%D1%96%D1%97.

13. Навіщо потрібні Powerline адаптери [Електронний ресурс] // Lantorg. – 2017. – Режим доступу до ресурсу: https://lantorg.com/article/zachem-nuzhny-powerline-adaptery.

14. Інтернет з розетки: загальні принципи роботи технології і огляд Powerline-адаптера TP-LINK TL-PA6010 [Електронний ресурс] // 3dNews. – 2014. – Режим доступу до ресурсу: https://3dnews.ru/821880.

15. Домашній міні-клімат-контроль своїми руками [Електронний ресурс]. – 2013. – Режим доступу до ресурсу: https://geektimes.ru/post/258012/.

16. Система "Розумний будинок" для заміського будинку на Arduino Mega2560, HC-05, SIM900, DHT11,3-x DS18B20, RTC-DS1302 [Електронний ресурс] // Arduino.ru. – 2015. – Режим доступу до ресурсу: http://arduino.ru/forum/proekty/sistema-umnyi-dom-dlya-zagorodnogo-doma-na-arduino-mega2560-hc-05-sim900dht113-kh-ds18.

17. Фролов А. В. Синтез и распознавание речи. Современные решения. [Електронний ресурс] / А. В. Фролов, Г. В. Фролов. – 2003. – Режим доступу до ресурсу: http://www.frolov-lib.ru/books/hi/index.html.

18. Квитко М.В. Распознаваниеречи с помощью глубокихрекуррентных нейронныхсетей [Електронний ресурс]/ Квитко М.В. // IASA –2016 р. –223 стр. –Режим доступу: http://sait.kpi.ua/media/filer_public/73/32/7332a68e-e93b-4c57-a3c8-66f11ee074cd/sait2016ebook.pdf

19. Голосове управління Arduino засобами Processing і Google Speech API [Електронний ресурс]. – 13. – Режим доступу до ресурсу: https://habrahabr.ru/post/236673/.

20. Мясищев А. А. Управление голосом с помощью Android и Arduino [Електронний ресурс] / А. А. Мясищев. – 2015. – Режим доступу до ресурсу: http://khnu.km.ua/root/kaf/ksm/my_syte_g/.

21. Mohri M. Speech recognition with weighted finite-state transducers. In Springer Handbook of Speech Processing / M. Mohri, M. Pereira, F. Riley. // Springer Berlin Heidelberg. – 2008. – С. 559–584.

22. Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N. et al. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. Signal Processing Magazine, IEEE, 29(6), 82-97.

23. Jurafsky D., Martin J.H. (2008) Speech and language processing, 2nd edition. Prentice Hall.

24. Голосове керування [Електронний ресурс] // Wikipedia. – 2007. – Режим доступу до ресурсу: https://ru.wikipedia.org/wiki/Голосовое_управление.

25. Дискретное преобразование Фурье. Википедия. [Electronic resource]. - Mode of access: https://ru.wikipedia.org/wiki/Дискретное_преобразование_Фурье.

26. Fast Artificial Neural Network Library(FANN). [Electronic resource]. - Mode of access: http://leenissen.dk/fann/wp/.

27. Audacity. Википедия. [Electronic resource]. - Mode of access: https://ru.wikipedia.org/wiki/Audacity.

28. Arduino. Официальный сайт. [Electronic resource]. - Mode of access: http://arduino.cc , 2015.

29. Круглов В. Искусственные нейронные сети / В. Круглов, В. Борисов. – Горячая Линия – Телеком, 2001.

30. Холоденко А.Б., “О построении статистических языковых моделей для систем распознавания русской речи” // Интеллектуальные системы, 2002. Т.6. Вип. 1-4. С. 381-394.

31. MIT Lectures 2003. http://ocw.mit.edu/courses/electrical-engineering-andcomputer-science/6-345-automatic-speech-recognition-spring-2003/downloadcourse-materials/

32. Фант. Г. Акустическая теория речеобразования. «Наука». Москва 1964. 4. Picone J. Fundamentals of speech recognition: a short course.1996. http://speech.tifr.res.in/tutorials/fundamentalOfASR_picone96.pdf

33. Алдошина И. Основы психоакустики. http://giga.kadva.ru/files/edu/AldoshinaPsychoacoustics.pdf

34. Слуховая система. серия "Основы современной физиологии". "Наука", Ленинград, 1990.

35. Seneff S. “Pitch and Spectral Analysis of Speech Based on an Auditory Synchrony Model”, Technical Report 504, January 1985 8. Hermansky H. (1997): “Should recognizers have ears?”, In RSR-1997, 1-10.

36. Маркел Дж.Д., Грей А.Х., Линейное предсказание речи, Москва,"Связь", 1980.

37. Hermansky H., Morgan N., "RASTA Processing of Speech", in IEEE Transaction on Speech and Audio Processing, Vol. 2, No. 4, pp. 587-589, October 1994.

38. .Карпов А.А., Кипяткова И.С., Методология оценивания работы систем автоматического распознавания речи // Известия вузов. Приборостроение, Т. 55, № 11, 2012, С. 38-43.

39. Левенштейн В.И., Двоичные коды с исправлением выпадений, вставок и замещений символов. Доклады Академий Наук СССР, 1965, 163.4:845- 848.

40. Kurimo M., Creutz M., Varjokallio M., Arsoy E., Saraclar M., Unsupervised segmentation of words into morphemes - Morpho challenge 2005

Application to automatic speech recognition. In Proc. INTERSPEECH-2006, Pittsburgh, USA, 2006, pp. 1021-1024.

41. Schlippe T., Ochs S., Schultz T., Grapheme-to-Phoneme Model Generation for Indo-European Languages. In Proc. ICASSP-2012, Kyoto, Japan, 2012.

42. Huang C., Chang E., Zhou J., Lee K. Accent modeling based on pronunciation dictionary adaptation for large vocabulary Mandarin speech recognition. In Proc. INTERSPEECH-2000, Beijing, China, 2000, pp. 818-821

43. Hannemann M., “Combinations of Confidence Measures for the Detection of Out-of-Vocabulary Segments in Large Vocabulary Continuous Speech Using Differently Constrained Recognizers”, Otto-von-Guericke-Universitat Magdeburg, 21. April 2008.

44. Bourlard H., Wellekens C.J., “Links between Markov models and multilayer perceptrons”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12 , No. 12, 1990, pp. 1167-1178.

45. Bourlard H., Hermansky H., Morgan N., “Towards increasing speech recognition error rates”, Speech Communication, Vol. 18, 1996, p.p. 205–231.

46. Hornik K., Stinchcombe M., White H., “Multilayer feedforward networks are universal approximators”, Neural Netw. Vol. 2(5), 1989, pp. 359–366.

47. Hinton G., Deng L., Yu D., Dahl G., Mohamed A., Jaitly N., Senior A., Vanhoucke V., Nguyen P., Sainath T., Kingsbury B., “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups”, IEEE Signal Process. Mag., Vol. 29, No. 6, Nov. 2012, pp. 82–97. 55. Dong Yu, Li Deng, “Automatic Speech Recognition. A Deep Learning Approach”, Springer-Verlag, London. 2015, 321 p.

48. Чистович Л.А. и др., «Руководство по физиологии. Физиология речи. Восприятие речи человеком», «Наука», Ленинград, 1976.

49. Hermansky H., Ellis D., Sharma S., “Tandem connectionist feature extraction for conventional HMM systems”, Proc. ICASSP-2000, Istanbul. 2000. V. 3. pp. 1635–1638.

50. Deng, L., Chen, J., “Sequence classification using high-level features extracted from deep neural networks.” In: Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014, pp. 6894-6898.

51. Hochreiter S., Schmidhuber J., “Long short-term memory.” Neural Computation, V. 9(8), 1997, pp. 1735–1780.

52. Pascanu R., Mikolov T., Bengio Y., “On the difficulty of training recurrent neural networks”, Cornell University Library, arXiv:1211.5063 [cs.LG], 2013.